A UNITED STATES DEPARTMENT OF COMMERCE PUBLICATION

SECTION B

JOURNAL OF RESEARCH

of the NATIONAL BUREAU OF STANDARDS

JNBBAU 76B (3&4) 91 – 220, July-Dec. 1972

Mathematical Sciences

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards

Journal of Research

of the

National Bureau of Standards

B. Mathematical Sciences

July–Dec. • 1972

VOLUME • 76B

NUMBERS • 3&4

Editor: Morris Newman

Associate Editor: F. W. Olver

Publication dates: Mar. 10, June 10, Sept. 10, Dec. 10, 1972

J. Res. Nat. Bur. Stand. (U.S.), B

JOURNAL OF RESEARCH

The National Bureau of Standards Journal of Research reports research and development in the fields of activity shown at right. Also included from time to time are survey articles on topics closely related to the Bureau's scientific and technical program.

The Journal is published in two separate sections as follows:

A. Physics and Chemistry

Contains papers of interest primarily to scientists working in these fields. Issued six times a year. Annual subscription: domestic, $17.00; foreign, $21.25. Single copy, varies.

B. Mathematical Sciences

Presents studies and compilations designed mainly for the mathematician and the theoretical physicist. Issued quarterly. Annual subscription: domestic, $9.00; foreign, $11.25. Single copy $2.25.

The National Bureau of Standards[1] was established by an act of Congress March 3, 1901. The Bureau's overall goal is to strengthen and advance the Nation's science and technology and facilitate their effective application for public benefit. To this end, the Bureau conducts research and provides: (1) a basis for the Nation's physical measurement system, (2) scientific and technological services for industry and government, (3) a technical basis for equity in trade, and (4) technical services to promote public safety. The Bureau consists of the Institute for Basic Standards, the Institute for Materials Research, the Institute for Applied Technology, the Center for Computer Sciences and Technology, and the Office for Information Programs.

THE INSTITUTE FOR BASIC STANDARDS provides the central basis within the United States of a complete and consistent system of physical measurement; coordinates that system with measurement systems of other nations; and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. The Institute consists of a Center for Radiation Research, an Office of Measurement Services and the following divisions:

Applied Mathematics — Electricity — Mechanics — Heat — Optical Physics — Linac Radiation[2] — Nuclear Radiation[2] — Applied Radiation[2] — Quantum Electronics[3] — Electromagnetics[3] — Time and Frequency[3] — Laboratory Astrophysics[3] — Cryogenics[3].

THE INSTITUTE FOR MATERIALS RESEARCH conducts materials research leading to improved methods of measurement, standards, and data on the properties of well-characterized materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government agencies; and develops, produces, and distributes standard reference materials. The Institute consists of the Office of Standard Reference Materials and the following divisions:

Analytical Chemistry—Polymers—Metallurgy—Inorganic Materials—Reactor Radiation—Physical Chemistry.

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services to promote the use of available technology and to facilitate technological innovation in industry and Government; cooperates with public and private organizations leading to the development of technological standards (including mandatory safety standards), codes and methods of test; and provides technical advice and services to Government agencies upon request. The Institute also monitors NBS engineering standards activities and provides liaison between NBS and national and international engineering standards bodies. The Institute consists of a Center for Building Technology and the following divisions and offices:

Engineering and Product Standards—Weights and Measures—Invention and Innovation—Product Evaluation Technology—Electronic Technology—Technical Analysis—Measurement Engineering—Building Standards and Code Services[4]—Housing Technology[4]—Federal Building Technology[4]—Structures, Materials and Life Safety[4]—Building Environment[4]—Technical Evaluation and Application[4]—Fire Technology.

THE INSTITUTE FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides technical services designed to aid Government agencies in improving cost effectiveness in the conduct of their programs through the selection, acquisition, and effective utilization of automatic data processing equipment; and serves as the principal focus within the executive branch for the development of Federal standards for automatic data processing equipment, techniques, and computer languages. The Center consists of the following offices and divisions:

Information Processing Standards—Computer Information—Computer Services—Systems Development—Information Processing Technology.

THE OFFICE FOR INFORMATION PROGRAMS promotes optimum dissemination and accessibility of scientific information generated within NBS and other agencies of the Federal Government; promotes the development of the National Standard Reference Data System and a system of information analysis centers dealing with the broader aspects of the National Measurement System; provides appropriate services to ensure that the NBS staff has optimum accessibility to the scientific information of the world, and directs the public information activities of the Bureau. The Office consists of the following organizational units:

Office of Standard Reference Data—Office of Technical Information and Publications—Library—Office of International Relations.

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.
[2] Part of the Center for Radiation Research.
[3] Located at Boulder, Colorado 80302.
[4] Part of the Center for Building Technology.

U.S. DEPARTMENT OF COMMERCE
Peter C. Peterson, Secretary

NATIONAL BUREAU OF STANDARDS
Lawrence M. Kushner Acting Director

Order all publications from the Superintendent of Documents
U.S. Government Printing Office, Washington, D.C. 20402.

The Secretary of Commerce has determined that the publication of this periodical is necessary in the transaction of the public business required by law of this Department. Use of funds for printing this periodical has been approved by the Director of the Office of Management and Budget through June 30, 1976.

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Vol. 76B, Nos. 3 and 4, July–December 1972

"Plural-Series" Approximations of Functions*

L. V. Spencer

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

(May 15, 1972)

A method is given for developing approximations to neutron and gamma-ray transport distributions in the form of a superposition of U_n^k polynomial series multiplied by exponentials and characterized by different scale factors. Convergence conditions and error bound expressions are given. Biorthogonality properties are worked out. Examples are given in which these approximations are compared with polynomial approximations.

Key words: Biorthogonal function; gamma-ray transport; moment methods; neutron transport; polynomial representations; radiation penetration.

1. Introduction

In the moment method of calculating space distributions in radiation transport theory, spatial moments defined by

$$M_n = \frac{1}{n!}\int_0^\infty dz z^n F(z), \tag{1}$$

or

$$M_n = \frac{1}{n!}\int_{-\infty}^\infty dz z^n F(z), \tag{1'}$$

are evaluated by numerical integration of a recursive system of Volterra integral equations of the second kind [1, 2].[1] For any of several reasons, one frequently encounters the problem of constructing distributions by use of alternate, rather than consecutive moments [2, 5]. Because spatial trends are basically exponential, this leads to the possibility of representations in a family of biorthonormal systems of polynomials $U_n^k(z)$, with adjoint polynomials $\hat{U}_n^k(z)$, satisfying the condition

$$\int_0^\infty dz z^k e^{-z} \hat{U}_n^k(z) U_m^k(z) = \delta_{nm}, \tag{2}$$

with

$$z^k \hat{U}_n^k(z) = \sum_{i=0}^n (-)^i \binom{n}{i} \frac{z^{k+2i}}{(k+2i)!}. \tag{2'}$$

Many other properties of these functions have been given, in references [1–4], and particularly in reference [4]. Perhaps we should say here that we draw freely on the results and terminology of reference [4] and reference [6] in this paper.

AMS Subject Classification: Primary 4110, Secondary 3036.

*Work supported by the Office of Naval Research and the Defense Civil Preparedness Agency.

[1] Figures in brackets indicate the literature references at the end of this paper.

The application of these polynomials to gamma-ray and neutron distributions leads to questions (a) of convergence in the limit, and (b) of convergence of a truncated series to within known error bounds. In the latter case, the utility of the method depends on the rapidity of convergence, because it is not practical to evaluate more than 6 to 10 of the series coefficients.

The problem of convergence in the limit was treated in reference [4], and a more recent paper [6] developed error bounds for truncated series. In addition, for gamma-ray distributions, the convergence does tend to be rapid. Thus the problem of constructing distributions from moments would appear to be completely solved. The convergence arguments are summarized in appendix B, with additions.

Unfortunately, however, many important problems do not have rapidly converging series. For example, if the radiation source has components of differing energy, their attenuation is a superposition of exponentials, and the more rapidly attenuated components give slowly converging contributions to the coefficients of the polynomial representation.

Another case is that in which the radiation initially travels at an oblique angle to the direction of penetration. A still more important case is that of the neutrons, where very high numbers of collisions with nuclei give spatial distributions which combine exponential with Gaussian features [7].

Existence of these problems led very early to use of nonlinear approximations which were designated "function-fitting" methods, in which the individual terms in an approximating series have the same functional form, with differing weight and differing scale factor [2]. While very useful, and in good agreement with the more basic polynomial representation, these methods have the limitation that it is difficult to assess their accuracy. In fact, it is easy to show that the more obvious error bounds which can be determined for such representations must be broader than error bounds for the polynomial representation. This is highly unsatisfactory, because one turns to these methods mainly when the polynomial representations are poor.

In this paper we develop a type of approximation which we designate a "plural-series" representation or approximation. It is a generalization of the polynomial approach which incorporates the flexibility of the function-fitting methods. But unlike the latter, one can prove convergence in the limit, and can evaluate bounds to truncation errors. Thus it combines the advantages of both methods. Because the polynomial method is a special case, we will refer to it herein as a "single-series" approximation.

We will show that it is possible to use plural-series calculations to evaluate error bounds for at least some function-fitting representations. This makes it possible to evaluate accuracy for previously published calculations.

We have begun to apply the plural-series method to types of distribution for which polynomial approximations converge poorly, and we include an example of this in which the distribution has no physical significance.

2. Plural-Series Representations

The representation which we wish to use in obtaining an approximation to $F(z)$ is

$$F_N(z) = \sum_{j=1}^{J} (A_j/x_j^{k+1}) F_N^j(z), \tag{3}$$

where

$$F_N^j(z) = \sum_{n=0}^{N-1} a_n^j U_n^k(z/x_j) e^{-z/x_j}. \tag{3'}$$

The x_j are arbitrarily chosen scale factors and the A_j are arbitrary multipliers. The approximate function $F_N(z)$ is required to have N moments which have values exactly agreeing with known values $M_k, M_{k+2}, M_{k+4}, \ldots$ for the corresponding sequence of moments of $F(z)$.

For $J > 1$, there are $J \cdot N$ coefficients a_n^j in (3), a much larger number than the number of moments, N. Hence other requirements must be invoked to determine the coefficients. For this purpose we use a simple variational rule, that the sum S_N defined as follows is to be minimized:

$$S_N = \sum_{j=1}^{J} (A_j / x_j^{k+1}) \sum_{n=0}^{N-1} (c_n a_n^j)^2, \tag{4}$$

where the c_n are arbitrary weighting coefficients. This leads to an elementary Lagrange multiplier calculation for the coefficients a_n^j, with side conditions as follows expressing the requirements of the moments of the representation:

$$M_{k+2n} = \sum_{j=1}^{J} (A_j / x_j^{k+1}) M_{k+2n}^j, \tag{5}$$

where

$$M_{k+2n}^j = \frac{1}{(k+2n)!} \int_0^\infty dz z^{k+2n} F_N^j(z). \tag{5'}$$

Because of the simplicity of (2'), the expression for a_n^j in terms of the moments of the "partial" functions F_N^j is very simple. Since

$$a_n^j = \int_0^\infty d(z/x_j)\,(z/x_j)^k \hat{U}_n^k(z/x_j) F_N^j(z), \tag{6}$$

insertion of the expression (2') for $\hat{U}_n^k$ into (6) gives

$$a_n^j = \sum_{i=0}^{n} (-)^i \binom{n}{i} M_{k+2i}^j / x_j^{k+2i+1}. \tag{7}$$

Before proceeding with the further analysis, let us mention restrictions to be observed on the various types of arbitrary coefficients. The x_j must be positive or else the integrals do not converge. In addition, we will limit our considerations to positive values of A_j so that S_N will always be positive. Then S_N turns out to be a squared norm and plays a central role in the discussion of error bounds. In addition, we assign positive values to the weighting coefficients c_n, which play a role here similar to that of the c_n coefficients in reference [6].

What the minimizing condition on S_N does is to select a particular biorthogonal system for the approximation. Our task is largely that of familiarizing ourselves with properties of this system, which turn out to be attractive as well as useful.

3. Evaluation of Coefficients

Equation (7) expresses a linear system whose formal solution makes it easier to express the main calculation to be performed. In matrix form, (7) is seen to be simply

$$x_j^{k+1} \begin{pmatrix} a_0^j \\ a_1^j \\ a_2^j \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot \\ 1 & -1 & 0 & \cdot & \cdot & \cdot \\ 1 & -2 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & x_j^{-2} & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & x_j^{-4} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} M_k^j \\ M_{k+2}^j \\ M_{k+4}^j \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}$$

Since the binomial matrix is its own inverse, we can immediately write

$$\begin{pmatrix} M_k^j \\ M_{k+2}^j \\ M_{k+4}^j \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} = x_j^{k+1} \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & x_j^2 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & x_j^4 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot \\ 1 & -1 & 0 & \cdot & \cdot & \cdot \\ 1 & -2 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} a_0^j \\ a_1^j \\ a_2^j \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}. \tag{8}$$

Designating the matrix elements of the product of the first two matrices on the right by γ_{in}^j, we have the desired result,

$$M_{k+2i}^j = x_j^{k+1} \sum_{n=0}^{N-1} \gamma_{in}^j a_n^j . \tag{9}$$

Let us note in passing that the γ_{in}^j are the coefficients which, according to (2′), change a series of $\hat{U}_n^k$ polynomials into a power series,

$$\frac{z^{2i}}{(k+2i)!} = \sum_{n=0}^{N-1} \gamma_{in}^j \hat{U}_n^k(z/x_j). \tag{10}$$

To obtain the form of the side conditions which we wish to use, we insert (9) into (5), with the result

$$M_{k+2i} = \sum_{j=1}^{J} A_j \sum_{n=0}^{N-1} \gamma_{in}^j a_n^j. \tag{11}$$

Then, designating the N Lagrange multipliers as $\lambda_1, \lambda_2, \ldots \lambda_N$, we write out the quantity to be minimized,

$$S_N = \sum_{j=1}^{J} (A_j/x_j^{k+1}) \left\{ \sum_{n=0}^{N-1} (c_n a_n^j)^2 \right\} + 2 \sum_{i=0}^{N-1} \lambda_i \left\{ M_{k+2i} - \sum_{j=1}^{J} A_j \sum_{n=0}^{N-1} \gamma_{in}^j a_n^j \right\}. \tag{12}$$

Vanishing of linear variations of S_N with respect to the λ_i gives back equations (11). The conditions $\{\partial S_N/\partial a_n^j = 0\}$ yield the following additional $J \cdot N$ equations:

$$a_n^j = x_j^{k+1} \sum_{i=0}^{N-1} \lambda_i \gamma_{in}^j / c_n^2. \tag{13}$$

Formal solution of the linear system of $J \cdot (N+1)$ equations requires only a straightforward matrix inversion. We insert (13) into (11) to obtain

$$M_{k+2i} = \sum_{i'=0}^{N-1} u_{ii'} \lambda_{i'}, \tag{14}$$

where

$$u_{ii'} = \sum_{j=1}^{J} A_j x_j^{k+1} \sum_{n=0}^{N-1} \gamma_{in}^j \gamma_{i'n}^j / c_n^2. \tag{15}$$

The matrix whose elements are $u_{ii'}$ is symmetric and readily evaluated. Writing $v_{nn'}$ for the elements of the inverse matrix, the equations (14) become

$$\lambda_n = \sum_{n'=0}^{N-1} v_{nn'} M_{k+2n'}, \tag{16}$$

and values of the a_n^j follow from (13).

4. Norm Convergence for $N \to \infty$

Let us assume that $J=1$, that x_1 and c_n values have been selected, and that for this case $\lim_{N\to\infty} S_N$ is finite. Then it follows that for $J>1$, for the same set of c_n values, and for a set of x_j values which include the same x_1, the sequence of S_N values must again have a finite limit. In addition, the sequence of S_N values is monotone increasing.

To demonstrate that these assertions hold true we first note that F_{N+1} is an approximation which fits the same moments as does F_N (plus one more). If, in the expression for F_{N+1}, one arbitrarily changes the a_N^j values to zero, for all j, the first N moments are unaffected due to biorthogonality. Call such a modified approximation F'_{N+1}, and calculate the corresponding quantity S'_{N+1}. Then

$$S_{N+1} \geq S'_{N+1} \geq S_N, \qquad J \geq 1. \tag{17}$$

The first of the inequalities holds because the a_N^j would not usually be zero, while the second holds because S_N was calculated to have the smallest value possible for a set of approximations which includes S'_{N+1}. Thus the S_N sequence is monotone increasing, as well as positive.

That the S_N sequence for $J>1$ is bounded follows if the $J=1$ sequence converges, and if any one of the x_j has the same value in the plural series case as the x_1 of the single-series case, as we have assumed. After all, for each value of N, the single-series approximation is only one in which $a_n^j=0$ for all j except that corresponding to the single-series x_1. It is therefore an element in the set of approximations among which the plural-series S_N is smallest. Hence the value of S_N, $J>1$, is always bounded above by the single-series limit and therefore the limit, $N\to\infty$, exists and is less than or equal to the single-series limit.

This argument makes no requirement on the values of x_j, $J>1$, other than convergence of a single-series sequence in which one of the x_j values is used as scale factor. Therefore, plural-series approximations can include components with exponential trends inappropriate for use in single-series representations.

5. Transformation Properties

The transformation properties of plural-series approximations follow from the fact that adjoint to $F_N(z)$ is a function which we designate $\hat{F}_N$,

$$\hat{F}_N(z) = \sum_{i=0}^{N-1} \lambda_i \frac{z^{2i}}{(k+2i)!}. \tag{18}$$

We establish this by an argument in terms of "transformation kernels", and for this purpose we first observe that

$$\{c_n^{-1} U_n^k(z) e^{-z}\} = \int_0^\infty dy y^k e^{-(z+y)} T^k(z,y) \{c_n \hat{U}_n^k(y)\}, \tag{19}$$

where

$$T^k(z,y) = \sum_{n=0}^{\infty} \{c_n^{-1} U_n^k(z)\}\{c_n^{-1} U_n^k(y)\}. \tag{19'}$$

The series (19′) is assumed either to converge for all y, z or else to diverge at $z=0$, $y=0$ in such a way that the integration (19) still leads to convergent results. For $c_n = 1$, T^k is a combination of modified Bessel functions, with a divergence at $y = z = 0$ [4]. Because of the factor y^k, this divergence is harmless in the integration.

If we insert (10) into (18), and then make use of (13), we obtain the expression

$$\hat{F}_N(z) = x_j^{-(k+1)} \sum_{n=0}^{N-1} a_n^j c_n^2 \hat{U}_n^k(z/x_j). \tag{20}$$

From (19) it then follows that

$$F_N^j(z) = \int_0^{\infty} dy y^k e^{-(z+y)/x_j} T^k(z/x_j, y/x_j) \hat{F}_N(z), \tag{21}$$

so that

$$F_N(z) = \int_0^{\infty} dy y^k \left\{ \sum_{j=1}^{J} (A_j/x_j^{k+1}) e^{-(z+y)/x_j} T^k(z/x_j, y/x_j) \right\} \hat{F}_N(y). \tag{22}$$

Thus, the transformation appropriate to the plural-series representation has a kernel which is a simple linear combination of the kernels for the component series taken separately.

To conclude this argument, we note that multiplication of (13) by $(c_n)^2 a_n^j$, followed by summation over n, yields

$$\sum_{n=0}^{N-1} (c_n a_n^j)^2 = \sum_{i=0}^{N-1} \lambda_i M_{k+2i}^j = \int_0^{\infty} dz z^k \hat{F}_N(z) F_N^j(z), \tag{23}$$

so that

$$S_N = \int_0^{\infty} dz z^k \hat{F}_N(z) F_N(z). \tag{24}$$

6. Biorthogonality Properties of Plural-Series Systems

The discussion of error bounds is perhaps simplest in terms of the biorthogonal functions appropriate to plural-series cases. Let us therefore derive expressions for the two sets of biorthogonal functions, which we designate $P_N(z)$ and $\hat{P}_N(z)$.

We consider first the sequence of moments $M_k = 1$, and $M_{k+2n} = 0$, for $n > 0$. The "function" which has these moments is irrelevant, though it must basically be a Dirac delta function located at $z = 0$. In any case, when this "function" is represented in any orthogonal or biorthogonal system, we expect all the terms to appear. The sequence of approximate representations corresponding to this moment system will be designated $G_N(z)$; and a corresponding sequence of adjoint functions, $\hat{G}_N(z)$, can be constructed by use of (18).

Two sequences of difference functions give the biorthogonal functions which we seek. For $N \geq 1$,

$$P_N(z) = \{G_{N+1} - G_N\}/d_N, \tag{25}$$

and

$$\hat{P}_N(z) = \{\hat{G}_{N+1} - \hat{G}_N\}/d_N, \tag{26}$$

where d_N is a normalizing constant still to be specified.

Since G_{N+1} and G_N have identical moments M_{k+2n}, for $n=0, 2, \ldots 2(N-1)$, it is clear that P_N is orthogonal to $\hat{P}_n$ for $n < N$, the latter being a linear combination of even powers:

$$\int_0^\infty dz z^k \hat{P}_n(z) P_N(z) = 0, \; n < N. \tag{27}$$

To prove that $\hat{P}_N(z)$ is also orthogonal to P_n, for $n < N$, we write

$$\int_0^\infty dz z^k \hat{P}_N P_n = \int_0^\infty dz z^k \int_0^\infty dy y^k \left\{ \sum_{j=1}^J (A_j/x_j^{k+1}) T^k(z/x_j, y/x_j) e^{-(y+z)/x_j} \right\} \hat{P}_N(z) \hat{P}_n(y)$$

$$= \int_0^\infty dy y^k \hat{P}_n(y) P_N(y) = 0. \tag{27'}$$

Therefore (25) and (26) describe a biorthogonal system which obeys the transformation (22). To complete the system we add the terms

$$P_0(z) = G_1(z)/d_0, \tag{28}$$

$$\hat{P}_0(z) = \hat{G}_1(z)/d_0. \tag{28'}$$

The normalizing condition for this system is

$$\int_0^\infty dz z^k \hat{P}_N(z) P_N(z) = 1, \tag{29}$$

which may be written by means of (24) and (4) in the form

$$d_N^2 = \sum_{j=1}^J (A_j/x_j^{k+1}) \sum_{n=0}^N c_n^2 \{ (a_n^j)_{N+1} - (a_n^j)_N \}^2, \tag{30}$$

$$d_0^2 = \sum_{j=1}^J (A_j/x_j^{k+1}) c_0^2 (a_0^j)_1^2, \tag{30'}$$

where $(a_n^j)_N$ is the coefficient a_n^j, determined for the construction G_N, and $(a_N^j)_N$ is always zero.

This completes the specification of the plural series biorthonormal system. In the special single-series case, there is cancellation between the quantities in braces in (30); and the usual normalization is recovered if we choose $c_n = 1$, and $A_1 = x_1^{k+1}$.

7. Convergence of Plural-Series Approximations

Norm convergence of plural-series approximations was earlier shown to hold if there is norm convergence of any of the single-series approximations which can be formed using the plural-series scale factors x_j. But this does not guarantee pointwise convergence of these approximations. To examine the latter, let us first note that (3) may be rewritten in the form of a P_N series,

$$F_N = \sum_{n=0}^{N-1} f_n P_n(z), \tag{31}$$

with

$$\|F_N\|^2 = \sum_{n=0}^{N-1} f_n^2 = S_N. \tag{32}$$

The form (31) follows from (22); one must recall that $\hat{F}_N$ is a polynomial and can therefore be written as a linear combination of the polynomials $\hat{P}_n$. Equation (32) follows from the normalization (29),

$$\|P_N\|^2 = 1. \tag{33}$$

We proceed by designating the kernel function of the transformation (22) as $\mathcal{T}^k(z, y)$, with the following extension of the notation to the N'th finite sum:

$$\overline{\mathcal{T}}_N^k(z, y) = \sum_{j=1}^{J} (A_j/x_j^{k+1}) \sum_{n=0}^{N-1} c_n^{-2} U_n^k(z/x_j) U_n^k(y/x_j) e^{-(y+z)/x_j}. \tag{34}$$

Next, applying (22) to $\hat{P}_n$, but noting that $U_{n'}^k(z/x_j)$ is orthogonal to the polynomial $\hat{P}_n(z)$ for all j and all $n' \geq N > n$, we observe that

$$P_n(y) = \int_0^\infty dz z^k \mathcal{T}^k(z, y) \hat{P}_n(z) = \int_0^\infty dz z^k \overline{\mathcal{T}}_N^k(z, y) \hat{P}_n(z). \tag{35}$$

The second equality holds for $n < N$, and due to biorthonormality, $\overline{\mathcal{T}}_N^k(z, y)$ must have the form

$$\overline{\mathcal{T}}^k(z, y) = \sum_{n=0}^{N-1} P_n(y) P_n(z) + \sum_{n=N}^{\infty} r_n(y) P_n(z), \tag{36}$$

where we do not have to concern ourselves with values for the coefficients $r_n(y)$ of the remainder term. From (36) and (32) it follows that for fixed y,

$$\sum_{n=0}^{N-1} [P_n(y)]^2 \leq \|\overline{\mathcal{T}}_N^k\|^2 < \mathcal{T}^k(y, y), \tag{37}$$

since $\mathcal{T}^k(y, y)$ is identical with $\|\mathcal{T}^k\|$.

Using this result together with the Schwarz inequality, for $z \neq 0$, we see that if the set of coefficients f_n in (31) and (32) has the property $S_N \leq S_\infty < \infty$, then

$$\left| \sum_{n=M}^{N} f_n P_n(z) \right|^2 \leq \left\{ \sum_{n=M}^{N} [P_n(z)]^2 \right\} \sum_{n=M}^{N} f_n^2 \leq \mathcal{T}^k(z, z) \sum_{n=M}^{N} f_n^2. \tag{38}$$

The term on the right goes to zero as $M, N \to \infty$; hence the Cauchy criterion is satisfied, and $\Sigma f_n P_n$ converges. Thus, pointwise convergence of a plural-series approximation follows if there is norm convergence of any of the single-series representations based on the plural-series scale factors x_j.

There remains the question as to whether the limit of a plural-series sequence of approximations agrees with that of the single-series approximations. To establish this we refer to the form (3), (3′), obtainable for any representation (31). We first note that if x_J is the largest of the plural-series scale factors, then the functions

$$F^j(z) = \sum_{n=0}^{\infty} a_n^j U_n^k(z/x_j) e^{-z/x_j}, \; j = 1, 2, \ldots, J, \tag{39}$$

in any convergent representation of the type (3), (3′), can also be represented by a convergent series using $U_n^k(z/x_J)$ only. One has, therefore, a convergent single-series representation for

$$\sum_{j=1}^{J} (A_j/x_j^{k+1})F^j(z), \tag{39'}$$

if one exists for $F(z)$. The coefficients of the single-series representation of this function must be exactly those of the corresponding single-series representation for $F(z)$, because both are given by the same linear combinations of the moments of $F(z)$ if the same series representation (based on x_J) is used for both. Hence (39′) is equivalent to a single-series representation of $F(z)$ in the limit.[2]

8. Bounds to Truncation Errors

To calculate bounds to truncation errors we require norm convergence (37), which follows from norm convergence of a single-series representation based on one of the x_j, as already stated.

As a result of (35), the kernel of the transformation (22) must be representable in a form analogous to (19′),

$$\mathcal{T}^k(z, y) = \sum_{j=1}^{J} (A_j/x_j^{k+1})e^{-(z+y)/x_j} T^k(z/x_j, y/x_j) = \sum_{n=0}^{\infty} P_n(z)P_n(y). \tag{40}$$

This follows from biorthogonality, plus the fact that, considered as a function of z, say, for fixed y, this function can be represented as a series of $U_n^k(z/x_J)$ functions with finite norm, where x_J is the largest scale factor. (See appendix B.)

Following the notation of reference [6], we put $y = z$ and make the designation

$$[D_0^k(z)]^2 = \sum_{j=1}^{J} (A_j/x_j^{k+1})e^{-2z/x_j} T^k(z/x_j, z/x_j) = \sum_{n=0}^{\infty} [P_n(z)]^2. \tag{41}$$

For approximations correct to N terms, one can accurately use the following truncated kernel function:

$$\sum_{n=0}^{N-1} P_n(z)P_n(y). \tag{42}$$

If we now apply the Schwartz inequality to the quantity $(F - F_n)$ as in (19′) of reference [6], we obtain

$$|F - F_N|^2 \leq \left\{\sum_{n=0}^{\infty} (f_n)^2\right\} \sum_{n=N}^{\infty} \{P_n(z)\}^2 \leq S_\infty\left\{[D_0^k(z)]^2 - \sum_{n=0}^{N-1} [P_n(z)]^2\right\}, \tag{43}$$

or

$$|F - F_N| \leq \sqrt{S_\infty}\, D_N^k(z), \tag{43'}$$

where we employ the notation of reference [6] except that exponentials now appear both in the function being approximated and in the definition of $D_N^k(z)$,

$$[D_N^k(z)]^2 = [D_0^k(z)]^2 - \sum_{n=0}^{N-1} [P_n(z)]^2. \tag{44}$$

We can follow reference [6] farther by defining an "error build-up" function $Eb(z)$,

[2] See Appendix B for a discussion of the automatic exclusion of functions having all moments zero.

$$Eb(z) = \sqrt{S_\infty}\, D_0^k(z)/F(z), \tag{45}$$

so that the (fractional) bounds to the truncation error are given by

$$\frac{|F - F_N|}{F} = Eb(z) \left\{\frac{D_N^k(z)}{D_0^k(z)}\right\}. \tag{46}$$

Both factors on the right of (46) reduce in the single-series case to the results of reference [6] because the extra exponential factor cancels out.

EXAMPLES: In transport problems, different energy components, characterized by exponentials which decrease at different rates, are commonly superposed. We choose a very simple case of this type as an example, namely

$$g(z) = 0.75e^{-z/0.2} + 0.25e^{-z}. \tag{47}$$

Table 1 gives the function $e^z g(z)$, together with five approximations, of which (a) is a single-series case, while (b) is a 3-term plural-series case. The exponential coefficients x_j are listed below the table of values.[3] The plural-series approximation (b) has a component with one scale factor very near unity and another which is about 0.33, and therefore much closer to the value 0.2 in (47). All calculations were performed with the computer program sketched in appendix A.

Below the scale factors in table 1 and norms S_N are listed for 5 through 8 terms in the approxi-

[3] See Appendix A for the rules used in selecting both x_j and A_j values.

TABLE 1. *Values of $g(z)e^z$ and five approximations.*
Scale factors x_j are listed below the values of the function. At the bottom of the table are the last four values of S_N together with an extrapolated value for S_∞.

z	Exact	(a) $k=0$ $J=1$	(b) $k=0$ $J=3$	(c) $k=2$ $J=3$	(d) $k=6$ $J=3$	(e) $k=6$ $J=3$
0	1.0000	0.680	0.799	1.000	1.000	1.000
0.05	0.8641	.650	.737	0.864	0.863	0.864
0.10	.7527	.622	.682	.753	.748	.753
0.2	.5870	.569	.588	.587	.574	.589
0.4	.4014	.478	.452	.402	.377	.406
0.6	.3180	.406	.366	.318	.292	.323
0.8	.2806	.349	.312	.280	.252	.284
1.0	.2637	.305	.278	.263	.246	.265
1.5	.2519	.241	.242	.252	.243	.249
2.0	.2503	.220	.237	.251	.248	.251
3.	.2500	.238	.248	.250	.254	.249
5.	.2500	.271	.256	.251	.248	.250
7.	.2500	.232	.243	.250	.251	.250
10.	.2500	.255	.254	.251	.250	.250
13.	.2500	.324	.267	.253	.249	.250
16.	.2500	.0835	.199	.241	.254	.250
20.	.2500	.0717	.241	.264	.243	.249
24.	.2500	3.505	1.08	.297	.251	.252
28.	.2500			.722	.297	.254
34.	.2500			−1.06	.077	.226
	x_1	1.000	0.991	0.801	0.850	0.771
	x_2		.572	.462	.491	.474
	x_3		.330	.267	.283	.280
S_N, $N=5$		0.234	0.2761	0.7506	0.7606	0.01013
6		.249	.2789	.7514	.7611	.01014
7		.262	.2821	.7521	.7611	.01023
8		.275	.2856	.7524	.7613	.01023
S_∞, extrapolated:		(0.446)	(0.324)	(0.759)	(0.763)	(0.0108)

mation. The final value in parentheses represents a very rough (linear) extrapolation to S_∞ using $N^{-1/2}$ as variable; while arbitrary, it tends to be quite conservative:

$$S_\infty \approx S_{N+1} - \frac{1}{\sqrt{N+1}} \left\{ \frac{S_{N+1} - S_{N-1}}{\frac{1}{\sqrt{N+1}} - \frac{1}{\sqrt{N-1}}} \right\} \tag{48}$$

The difference between the 8th norm and the extrapolation is an indicator of the degree of convergence, although the extrapolated value cannot be taken very seriously, particularly when, as in case (b), successive differences are still increasing. It turned out that with increasing N, these extrapolations first increased, and then in high-quality approximations decreased to meet the (rising) norm values.

Both the approximation values and the data on norms indicate that the plural-series approximation is better, the single-series approximation being generally unacceptable.

The remaining three columns represent an attempt to improve the approximations through use of information about the value and the first derivative of $g(z)$ at $z = 0$. This information makes possible the approximation of a modified function, selected to be

$$g'(z) = g(z) - e^{-z/0.25} \tag{49}$$

which has a trend near $z = 0$ which is proportional to z^2, and which can thus be represented with plural-series approximations involving U_n^k polynomials with $k \leq 2$. Three-term plural-series approximations are used in all cases (*c*)–(*e*). The cases (*d*) and (*e*) ignore the first four even moments and use U_n^6 polynomials [9]. In addition, for case (*e*) the x_j values were scaled down according to the rule

$$x_j' = \frac{x_j}{\sqrt{1 + a x_j^2}}, \quad a = 0.3. \tag{50}$$

This form is suggested by a condition that a single adjoint function of the form $\cos(az)$ be transformed by the appropriate kernel functions $T(z/x_j', y/x_j')$, as in (21), to the exponentials e^{-z/x_j}.

Some of the sharp behavior for z near zero is subtracted out by the additional term in (49), so that one expects improvement in the convergence, in the values for large z, and especially for small z which is now given correctly in the limit. But because there is sharp behavior remaining, one does not expect the improvement to be as dramatic as shown.

We should note that reducing the scale factor would also improve (*a*) and (*b*); the advantages in rising build-up factors which were discussed in references [6] and [8] apply here also.

Dropping the lower moments tends to degrade the approximation (*d*) relative to the approximation (*c*) for small z values, while introducing an improvement for large z values. Reducing the x_j values, on the other hand, improves the quality at all z. Recall that at least one of the x_j must always give a converging single-series approximation. That is the case for all approximations here shown.

Table 2 compares errors with estimated error bounds.[4] In all cases shown, we have used S_8 rather than the extrapolated approximation to S_∞ as multiplier in estimating error bounds. As a result, in (*a*) at $z = 0.4$ the estimated error bound falls slightly below the error. This shows the problem of "estimating" an error bound in a poorly convergent case. Perhaps the extrapolated S_∞ would be a better multiplier to use generally, though there is not much difference when convergence is at all satisfactory.

Elsewhere the estimated error bounds are reasonable, even excessive. Some use of the option $c_n \neq 1$ was made, with the following prescription [6]:

$$c_n = (n+1)^{m/4}, \tag{51}$$

[4] See appendix A for details on the calculations. They made use of (44) to (46), with data on D_0^k obtained using (19') with $y = z$.

TABLE 2. *Comparisons between the fractional error,* $(F-F_N)/F$, *designated "Err," and an estimated fractional error bound, designated "EB."*
Read "A−B" as A×10^{-B}.

z	(a) $k=0$ $J=1$		(b) $k=0$ $J=3$		(c) $k=2$ $J=3$		(d) $k=6$ $J=3$		(e) $k=6$ $J=3$	
	Err	EB	Err	EB	Err	EB	Err	EB	Err	EB
0.	−0.32	0.50	−0.20	1.0	0.	0.	0.	0.	0.	0.
.05	−.25	.38	−.15	0.54	.2−4	.019	−0.0016	9.9	.25−3	1.2
.10	−.17	.27	−.094	.32	.8−4	.051	−.0061	28.	.99−3	3.6
.20	−.031	.11	0.001	.24	.3−3	.10	−.022	55.	.38−2	7.7
.40	0.19	.18	.13	.35	.6−3	.13	−.061	43.	.012	9.1
.6	.28	.29	.15	.39	−0.2−3	.16	−.083	25.	.016	7.7
.8	.24	.31	.11	.34	−.14−2	.25	−.082	16.	.012	5.5
1.0	.16	.28	.053	.25	−.21−2	.30	−.068	9.8	.44−2	3.4
1.5	−0.043	.17	−.041	.16	0.3−3	.22	−.034	3.3	−0.010	0.89
2.0	−.12	.23	−.054	.20	.3−2	.19	−.009	1.8	−.77−2	.68
3.0	−.049	.22	−.006	.12	−0.17−2	.23	0.016	0.84	0.52−2	.21
5.	0.083	.23	0.023	.15	0.28−2	.26	−.009	.36	−0.30−2	.087
7.	−0.071	.22	−.027	.14	−0.47−3	.26	0.003	.21	.13−2	.059
10.	0.019	.44	0.017	.22	0.6−4	.37	−.002	.20	−0.12−2	.044
13.	.30	1.2	.067	.73	.011	.88	0.004	.31	0.35−3	.037
16.	−0.67	7.7	−0.21	1.8	−0.037	2.4	.014	.64	.19−2	.052
20.							−0.027	1.2	−0.48−2	.12
24.							0.005	2.9	0.68−2	.21
28.							.19	11.6	.014	.45
34.									−0.096	2.5

with $m=0$ for cases (*a*) and (*b*), and $m=2, 3, 6$ for cases (*c*), (*d*), and (*e*), respectively. Large values of m can lead to poor convergence of the norms, but for the functions $g(z)$ and $g'(z)$ approximated here, choice of m does not affect ultimate convergence. The evidence of table 1 is that the choices were reasonable.

Unlike the single-series case, the value of m affects the approximation as well as the error bound functions.

Error bounds generally tend to become large as one approaches large values of z at which convergence breaks down. But one also finds extremely large error bounds for small values of z in cases (*d*) and (*e*). These are due to the insensitivity of large moments to the details of the distribution at small z. That these bounds are realistic was demonstrated in some calculations not shown, in which for some m values, with calculations otherwise like (*d*) and (*e*), the errors in the small z region were very large indeed. In this region, for large k, the approximations were either very good or very bad. Use of large k approximations should evidently be accompanied by supplementary studies at small z.

9. Additional Comments

It was mentioned earlier that the plural-series approach can be used to gauge the accuracy of representations of the function-fitting type which make use of sums of exponentials, as in eq (A4) of appendix A. This requires that the x_j be assigned the values of the specified representation, and that the plural-series solution have $a_n^j=0$ for $n>0$. The latter requirement may or may not be satisfied, depending on the norm of the function-fitting representation in the plural-series system. Other plural-series representations can have smaller norms, particularly when some of the A_j have opposite signs and are large.

When all the constant factors in expressions such as (A4), in appendix A, have the same sign, the simple sum of exponentials tends to be the plural-series solution for all N less than or equal to the number of moments given correctly by the approximation. Very likely, with properly stated conditions one could prove a theorem to this effect. But when this simple property does not hold, one can still modify norm values by adjusting the shape of the error function D_0^k through choice of the A_j values. Further, one can modify the moments of the unknown function by addition of

moments of known terms, with final subtraction of these functions from the plural-series construction. A simple case of this would be transferral of some of the negative terms in a function-fitting representation to the other side, with corresponding changes in the moments of the function to be constructed. But such a simple prescription cannot be recommended as a general procedure because it removes exponential components somewhat indiscriminately from the plural-series approximating system.

Selection of A_j values for the examples of the preceding section was based on the assumption that the error function D_o^k should have terms corresponding to an elementary function-fitting rule, as discussed in appendix A. Cases with peculiar features, such as large terms which nearly cancel, probably require more care in specification of the A_j values. But the quality of the approximations usually appears to be surprisingly independent of such assignments.

10. Appendix A. A Plural-Series Computer Program

A general-purpose computer program which we have called FFPOL was constructed to perform calculations such as those described as examples. The following operations are performed by the main constituent components:

(1) Specification of exponential scale factors x_j, by subroutines designated BCOF and SCALE.

(2) *Solution of the basic Lagrange multiplier problem*, using a routine called SEP.

(3) *Computation of the error estimate functions* D_N^k and D_o^k, with a routine ESEP, which makes use of SEP.

(4) *Evaluation of both approximation and error bound data*, with a subroutine designated UNESEP.

More specifically, the following are the computations performed by the main subroutines just mentioned:

The scale factors x_j are evaluated as a product

$$x_j = \xi_j / s, \tag{A1}$$

where the ξ_j are pre-specified according to the rule

$$\xi_j^2 = \beta^{j-1}, j = 1,2, \ldots, J, \tag{A2}$$

$$\xi_J = J^{-1}, \tag{A3}$$

which ensures that these values cover the unit interval in steps which are basically geometric, but with roughly uniform coverage. BCOF evaluated the ξ_j. The overall scale factor s is evaluated by SCALE so that an expression of the type

$$\sum_{j=1}^{J} (f_j / x_j^{k+1}) e^{-z/x_j} \tag{A4}$$

can correctly fit $(J+1)$ moments. This calculation also determines the A_j of (3) and (4), by an arbitrary rule such as $A_j = |f_j|$.[5]

The procedure followed by SEP in evaluating both the λ_i values and the a_n^j coefficients is a straightforward evaluation first of the γ_{in}^j matrix elements and then the $u_{ii'}$ matrix elements.

---

[5] The effect of this selection on the quality of the representation is not clear, and not as strong as one might expect. According to (41), specification of large A_j values for small x_j components would appear to emphasize the small z region in the D^k functions, for example. The data of tables 1 and 2 were obtained with the function-fitting parameters, as in (A4), in the combination $|f_j/x_j^{k+1}|$. Perhaps a more logical choice for the A_j of (3), (4) might be the combination f_j^2/x_j^{k+1}, which tends to give $D^k \propto |f_j|/x_j^{k+1}$, at least for the dominant terms.

The inversion of the latter matrix, followed by the matrix multiplications (16) and (13) completes the calculation.

The error estimate function D_0^k is evaluated by ESEP with (25), using (19′). The sums are carried to 30–40 terms ordinarily, with evaluation of the $U_n^k(z)$ mostly by recursion. The recursion formula,

$$4(n+1)(n+2)[U_{n+2}^k - U_{n+1}^k] = z^2 U_n^k + 2(n+1)(4n+2k+1)[U_{n+1}^k - U_n^k]$$
$$- (2n+k)(2n+k-1)[U_n^k - U_{n-1}^k], \qquad \text{(A5)}$$

was derived from the differential equation, using eq (17.16′), both given on p. 744 of reference [2]. The recursion approach eventually breaks down, particularly for large values of k; but with suitable restrictions, and depending somewhat on the size of the argument z, 30–40 terms gives accurate results. ESEP evaluates D_N^k by subtracting from $(D_0^k)^2$ terms of the type $P_n(z)^2$, as shown in (44).

Evaluation of the approximation data by UNESEP utilizes (3). The (estimated) error bound data is calculated using D_N^k data from ESEP, with S_N taken to be an adequate approximation to S_∞.

Finally, the master routine, FFPOL, first calls BCOF and SCALE. This is followed by an optional provision to modify the x_j values just determined to obtain a related set which does *not* have the property of fitting $(J+1)$ moments, through use of (50). Lastly, UNESEP is called to give both approximation and (estimated) error bound data. All calculations have used double precision (60 bits).

11. Appendix B. Convergence of U_n^k Polynomial Representations

We summarize and complete here the arguments of reference [4] relating to convergence of the U_n^k series representations.

(1) First, let us note two properties of these biorthogonal functions:

$$U_n^k(z) = O(n^{-1/2}), \text{ for } z \neq 0, \qquad \text{(B1)}$$

and, also for $z \neq 0$,[6]

$$\left(\frac{z}{|z|}\right)^k \left(\frac{1}{\sqrt{1-u}}\right)^{k+1} e^{-\frac{|z|}{\sqrt{1-u}}} = \sum_{n=0}^{\infty} u^n U_n^k(z) e^{-|z|}. \qquad \text{(B2)}$$

Note that (B1) does not usually hold at $z=0$, where $U_n^k(0) = O(n^{(k-1)/2})$. The U_n^k functions could be (but have not been) defined as coefficients of the different powers of u in the expansion (B2). This expansion converges uniformly for $z \neq 0$ on the unit circle $|u| \leq 1$, except that, given $\epsilon > 0$, we must have $|u-1| \geq \epsilon$.

(2) Next, let us consider an *even* function $F(z)$, $-\infty < z < \infty$, whose Fourier transform has singularities *only* to the right and far left of the equilateral hyperbola shown in figure B1, a pair of regions which we refer to as "I". The Fourier transform of $F(z)$, which we designate $\Phi(\eta)$, will be an even function of η and will actually have twin singularities symmetrically located about the imaginary axis. The Fourier inversion integral is

$$F(z) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} d\eta e^{-\eta|z|} \Phi(\eta). \qquad \text{(B3)}$$

---

[6] In physical problems, singular behavior at $z=0$ will be known and usually treated by a special calculation. The remainder at $z=0$ will often, though not always, vanish. Thus, treatment of this point as a special case does not lead to any practical difficulties. Incidentally, eq (40) of reference [4] which is the starting point for the arguments leading to this statement on convergence, can be derived by induction without reference to other expressions for U_n^k which include singular components at $z=0$. This development makes use of (38) of reference [4], and the commutator of the two operators s and $(\partial/\partial s)^k$ in the integrand of (40) of reference [4].

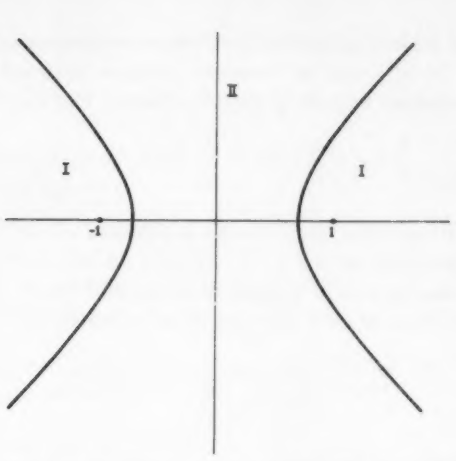


FIGURE B1. *The equilateral hyperbola satisfying* $|\eta^2-1|=|\eta^2|$, *which intersects the real axis at* $\pm 1/\sqrt{2}$. The Roman numerals identify Regions I and II.

Because of the location of the singularities, we can move the path of integration to a curve C completely to the right of the right-hand branch of the hyperbola.

To apply (B2) to this Fourier integral, we make a change of variables $u=(1-\eta^{-2})$, and note that the path of integration C can then be divided into a part C_1 located in the region of uniform convergence and a part C_2 corresponding to $|1-u|<\epsilon$. In the case of C_2, our change of variables gives

$$|\eta|>1/\sqrt{\epsilon}.$$

The two curve segments give two integrals; and for C_1 it is appropriate to use the series representation:

$$F(z)=\frac{1}{2\pi i}\int_{C_1}\frac{d\eta}{\eta}\,\Phi(\eta)\sum_{n=0}^{\infty}\left(\frac{\eta^2-1}{\eta^2}\right)^n U_n^0(z)e^{-|z|}+\frac{1}{2\pi i}\int_{C_2}\frac{d\eta}{\eta}\,e^{-\eta|z|}\,\Phi(\eta). \tag{B4}$$

The exponential factor in the integrand of the last term on the right has a magnitude determined by the real part of the exponent, and therefore at least as small as

$$\exp\left\{-\frac{|z|}{\sqrt{2\epsilon}}\right\}.$$

For $z\neq 0$, this factor can be made as small as one pleases; and we choose it so that the second term on the right of (B4) can be neglected altogether.

Because of uniform convergence, the integration and summation can be interchanged in the first term on the right of (B4), with the result

$$F(z)=\sum_{n=0}^{\infty}F_n U_n^0(z)e^{-|z|}, \tag{B5}$$

where

$$F_n=\frac{1}{2\pi i}\int_C\frac{d\eta}{\eta}\left(\frac{\eta^2-1}{\eta^2}\right)^n\Phi(\eta). \tag{B6}$$

We thus find that $F(z)$ has pointwise convergent representations of the type (B5). Also, by writing $\Phi(\eta)$ as a transform over $F(z)$, in (B6), we readily change to the more familiar form

$$F_n=\int_0^{\infty}dz\hat{U}_n^0(z)F(z). \tag{B6$'$}$$

(3) A parallel argument, also making use of the region of uniform convergence of (B2), but with a transformation $-u=\eta^2/(1-\eta^2)$ and an inversion integral confined to Region II, leads to the series form for the transformation kernels of the U_n^k systems: For $y, z>0$,

$$\left\{\frac{\partial^2}{\partial y \partial z}\right\}^k \frac{2}{\pi} K_0(\sqrt{y^2+z^2}) = \sum_{n=0}^{\infty} \{U_n^k(y)e^{-y}\}\{U_n^k(z)e^{-z}\}, \tag{B7}$$

where K_0 is the modified Bessel function. For the arguments which follow, the important feature of this expression is the fact that for $y=z>0$, the sum on the right converges to a finite value.

(4) Let some of the singularities of a function $\Phi(\eta)$ be located in Region II. Then it *may* be possible to choose a scale factor β such that the singularities of $\Phi(\eta/\beta)$ all lie in Regions I. If so, since

$$F(z)=\frac{\beta^{-1}}{2\pi i}\int_{-i\infty}^{i\infty} d\eta e^{-\eta|z|/\beta}\Phi(\eta/\beta), \tag{B8}$$

it is clear that a representation of the type

$$F(z)=\sum_{n=0}^{\infty} F_n' U_n^0(z/\beta)e^{-|z|/\beta} \tag{B9}$$

exists. But this will *not* be the case if any singularities of $\Phi(\eta)$ lie *on* the imaginary axis or if any point of the imaginary axis is a limit point for singularities of $\Phi(\eta)$.

(5) Preceding arguments applying only to the U_n^0 system can be generalized to the U_n^k systems, with $k>0$. If $F(z)$ is even and is known to have the following trend for small z values:

$$F(z)=z^k f(z), \tag{B10}$$

then $f(z)$ may be either even or odd. If the singularities of the Fourier transform φ of f lie entirely in Regions I for some scale factor β, the argument already given for the $k=0$ case leads to pointwise convergent series representations of the form

$$f(z)=\sum_{n=0}^{\infty} f_n^k U_n^k(z/\beta)e^{-|z|/\beta}, \tag{B11}$$

with

$$f_n^k=\frac{\beta^{-1}}{2\pi i}\int_C \frac{d\eta}{\eta^{k+1}}\left(\frac{\eta^2-1}{\eta^2}\right)^n \varphi(\eta/\beta)=\int_0^{\infty}\frac{dz}{\beta}(z/\beta)^k \hat{U}_n^k(z/\beta)f(z). \tag{B11'}$$

Equation (B11') makes it clear that $f(z)$ can contain singular terms at $z=0-$ Dirac delta functions and their derivatives – which do not contribute to the coefficients f_n^k because they are overriden by the factor z^k.

(6) Consider a function $F(z)$ with Fourier transform having singularities confined to Regions I. This same property will apply also to $f(z)$, if $F(z)$ has the form (B10). If we write expressions for finite approximations to $f(z)$,

$$f_N(z)=\sum_{n=0}^{N-1} f_n^k U_n^k(z)e^{-|z|} \tag{B12}$$

and apply Schwarz's inequality to the difference function $f-f_N$, we obtain the result

$$|f-f_N|^2 \leq \left\{\sum_{n=0}^{\infty}(f_n^k)^2\right\}\sum_{n=N}^{\infty}\{U_n^k(z)e^{-|z|}\}^2. \tag{B13}$$

Since the last factor on the right tends to vanish in the limit $N \to \infty$, it is clear that finite norm,

$$\sum_{n=0}^{\infty} (f_n^k)^2 < \infty,$$

implies pointwise convergence, for $z \neq 0$.

(7) Let $F(z)$ be even and representable by a U_n^0 series,

$$F(z) = \sum_{n=0}^{\infty} F_n U_n^0(z) e^{-|z|}. \tag{B15}$$

We further assume that this series has finite norm, and calculate the Fourier transform of both sides of (B15), by a limiting process:

$$\begin{aligned}\Phi(\eta) &= \lim_{\epsilon \to 0} \left\{ \int_{-\infty}^{-\epsilon} dz e^{\eta z} F(z) + \int_{\epsilon}^{\infty} dz e^{\eta z} F(z) \right\} \\ &= \lim_{\epsilon \to 0} \sum_{n=0}^{\infty} F_n \left\{ \int_{-\infty}^{-\epsilon} dz e^{\eta z} U_n^0(z) e^{-|z|} + \int_{\epsilon}^{\infty} dz e^{\eta z} U_n^0(z) e^{-|z|} \right\} \\ &= \sum_{n=0}^{\infty} F_n \frac{2}{1-\eta^2} \left(\frac{\eta^2}{\eta^2 - 1} \right)^n. \end{aligned} \tag{B16}$$

The limit $\epsilon \to 0$ can be taken as above providing $F(z)$ is sufficiently well-behaved near $z = 0$ that the Fourier transform exists.

Now, applying Schwarz's inequality to (B16) we obtain

$$|\Phi(\eta)| \leq \frac{2}{|\eta^2|} \|F\| \left\{ \sum_{n=0}^{\infty} \left| \frac{\eta^2}{\eta^2 - 1} \right|^{2n+2} \right\}^{1/2} \tag{B17}$$

And since $|\eta^2/(\eta^2 - 1)| < 1$, for η in Region II, the right side is finite at all points η in Region II. We conclude that for series representations with finite norm, the Fourier transform can have no divergent singularities in Region II.

(8) Note that for $z \neq 0$, the cases $-\infty < |z| < \infty$ and $0 < z < \infty$ can be considered equivalent, for functions $F(|z|)$ and $F(z)$, respectively. Hence problems in which the functions are defined on the infinite interval, can be analyzed with the use of U_n^k representations defined on the semi-infinite interval.

(9) Although we do not attempt it here, the existence of simple generating functions for the $\hat{U}_n^k$ polynomials makes a similar set of arguments possible, but in connection with even or odd functions whose *Laplace* transforms have singularities confined to Region II of figure B1.

(10) Physical solutions to neutron and gamma ray transport problems are such that the singularities of the Fourier transform lie on the real axis at distances from the origin $|\eta| \geq \mu_{\min}$, where $\mu_{\min}$ is the attenuation coefficient of the most penetrating radiation component and has a known value in any given problem. Thus one always expects convergent representations of the general type (B11) to exist.

One should realize that there exist functions of $|z|$ on the infinite interval whose moments *all* vanish. Such functions cannot be represented by U_n^k expansions; and one infers that these functions have Fourier transforms which are singular on the imaginary axis. This observation is confirmed by the discussion of these functions in reference [10]. Since there is neither physical nor mathematical reason to expect that such functions play any role in transport theory, their automatic exclusion from any representation based on U_n^k functions is appropriate.

479-206 O - 73 - 2

12. References

[1] Spencer, L. V., and Fano, U., Penetration and diffusion of x-rays. Calculation of spatial distributions by polynomial expansion, J. Research NBS 46, 446–456 (June 1951) RP2213.

[2] Fano, U., Spencer, L. V., and Berger, M. J., Penetration and diffusion of x-rays, Handbuch der Physik V. 38 (Springer-Verlag, Berlin, 1959).

[3] Preiser, S., An investigation of biorthogonal polynomials derivable from ordinary differential equations of the third order, J. Math. Anal. Appl. 4, No. 1, 38 (Feb. 1962).

[4] Spencer, L. V., and Flusser, P., Properties of a useful biorthogonal system, J. Res. Nat. Bur. Stand. (U.S.) 71B (Math. and Math. Phys.), No. 4, 197–211 (Oct.–Dec. 1967).

[5] Goldstein, H., and Wilkins, J. E., Jr., Calculation of the Penetration of Gamma Rays, Nucl. Development Associates, Inc. NYO–3075 (1954).

[6] Spencer, L. V., Bounds to Truncation Errors in Biorthogonal Polynomial Approximations, with Illustrative Applications to Gamma Ray Transport Distributions, J. Res. Nat. Bur. Stand. 75B (Math. Sci.), Nos. 1 and 2, 1–16 (Jan.–June 1971).

[7] Soran, P., and Goldstein, H., Reconstruction of neutron spatial distributions from spatial moments, Transactions of the American Nuclear Society 13, No. 1, 405 (1970).

[8] Spencer, L. V., and Simmons, G. L., Improved Moment Method Calculations of Gamma Ray Transport: Application to Point Isotropic Sources in Water, Nucl. Sci. Engr. 50, 20–30 (1973).

[9] Morris, E. E., An Extension of the Polynomial Approximation in Moments Method Calculations to the Deep Penetration of Gamma Radiation, Transactions 11, No. 1, 418–419 (1968).

[10] Shohat, J. A., and Tamarkin, J. D., The Problem of Moments, Mathematical Surveys No. 1, pp. 20, 22, American Mathematical Society (1943).

(Paper 76B3&4–364)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Vol. 76B, Nos. 3 and 4, July–December 1972

Scheduling a Time-Shared Server to Minimize Aggregate Delay*

A. J. Goldman

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

(July 21, 1972)

A simplified analysis is given of a problem situation, previously treated in the literature, which pertains to the delay-minimizing allocation of servicing times among N incoming streams requiring "processing" of some kind by a single "server" (e.g., a time-shared computer). The original problem is generalized to permit different "weights" for the delays suffered by different streams.

Key words: Computer systems; optimization; scheduling; time-sharing; traffic control.

1. Introduction

A 1967 paper by Rangarajan and Oliver [1][1] contains a formulation and analysis of the two problems described below, which pertain to the allocation of servicing times among N incoming streams requiring "processing" of some kind by a single "server." The server might for example be a switching point or a congestion point (e.g., a tunnel entrance) in a transport network, in which case "processing" an item (vehicle) simply means letting it through. Or, the server might be a computer handling reservations from several ticket offices, or exercising real-time control over vehicle movements on several network links, or performing some other tasks on a time-shared basis.

The streams are treated as continuous flows. During each service *cycle*, of duration T, the server handles stream 1 for time G_1, switches (with associated known switch-over or "dead" time) to handle stream 2 for time G_2, etc. The arrivals in each stream are assumed nonrandom, with a known uniform rate (possibly different for different streams). The server's processing rate, when serving a particular stream, is also assumed nonrandom and constant (possibly different for different streams). Each G_i is constrained to be at least large enough so that no queue remains in the ith stream when one of that stream's service periods ends.

The two problems formulated and analyzed are these:

PROBLEM 1: For given cycle time T, what allocation $G_1, G_2, \ldots, G_N$ of service times among the various streams is optimal, in the sense of minimizing total waiting time per cycle?

PROBLEM 2: What value of the cycle time T will minimize average waiting time?

Subsequently Horn [2] showed that the more general case, in which all streams are served equally often (possibly *more* than once) per cycle, can be reduced to PROBLEM 1. This provides additional reason for offering an alternative analysis, which is more self-contained and (at least to the writer) simpler than that of reference [1]. In addition, a mild generalization will be introduced by permitting the penalties for delay to be different for different streams.

AMS Subject Classification Number: Primary 9310.

* Research supported by the Northeast Corridor Transportation Project (Dept. of Transportation) and the NBSCenter for Computer Science and Technology.

[1] Figures in brackets indicate the literature references at the end of this paper.

2. Formulation

The data for PROBLEM 1 are

T = cycle time,
N = number of streams,
a_i = arrival rate for ith stream,
s_i = service rate limit when processing ith stream ($s_i > a_i$),
d_i = dead time in switching from ith stream to next one,
p_i = penalty factor for delays to ith stream.

Note that our (a_i, s_i, d_i) are the $(\lambda_i, \mu_i, \tau_i)$ of reference [1], which in effect assumes all $p_i = 1$.

Under the assumption of first-in-first-out service within each stream, the waiting time per cycle for the ith stream is found as in reference [1] to be

$$W_i[T, G_i] = \{a_i(T-G_i)^2 + a_i^2(T-G_i)^2/(s_i-a_i)\}/2$$

$$= a_i s_i (T-G_i)^2/2(s_i-a_i). \tag{1}$$

(The factor 1/2 was omitted from the analogous equation in reference [1]; also our (1) differs from that formula by a factor T because we work with total rather than time-averaged delay.) Thus the function to be minimized is

$$W^0 = \sum_1^N p_i a_i s_i (T-G_i)^2/2(s_i-a_i). \tag{2}$$

The condition, that each stream have its queue disappear before its service period ends, is expressed by

$$G_i(s_i-a_i) \geq (T-G_i)a_i \qquad \text{or equivalently} \qquad G_i s_i \geq T a_i,$$

which is equivalent to

$$(T-G_i)s_i \leq T(s_i-a_i). \tag{3}$$

The remaining constraint on the G_i's is the obvious identity which can be expressed, in terms of total service time and total dead time

$$G = \sum_1^N G_i \qquad \text{and} \qquad D = \sum_1^N d_i,$$

in the form

$$G + D = T. \tag{4}$$

We simplify by introducing the new variables

$$x_i = (T-G_i)/T,$$

and also

$$W = 2W^0/T^2 \tag{5}$$

as the new minimand, equivalent to the previous one since T is fixed for PROBLEM 1. Furthermore, let

$$b_i = (s_i-a_i)/s_i > 0,$$

$$c_i = p_i a_i/b_i = p_i a_i s_i/(s_i-a_i) > 0,$$

$$q_i = 1/c_i = b_i/p_i a_i > 0,$$

$$B = N - 1 + (D/T).$$

Then from (2) and (5), we see that PROBLEM 1 requires the minimization of

$$W(T) = W = \sum_1^N c_i x_i^2 \tag{6}$$

subject to the conditions (3), which are equivalent to

$$0 \leq x_i \leq b_i, \tag{7}$$

and to condition (4), which is equivalent to

$$\sum_1^N x_i = B. \tag{8}$$

From (7) and (8) we obtain the condition

$$B \leq \sum_1^N b_i, \tag{9}$$

which is both necessary and sufficient for the consistency of the constraints, and is assumed to hold in what follows.

3. Solution of PROBLEM 1

Since the problem requires minimizing a continuous strictly convex function over the closed bounded subset of x-space defined by (7) and (8), there must exist a unique relative minimum which is in fact the unique absolute minimum. Hence we need only derive enough necessary conditions, for a local minimum, to single out just one point in x-space.

The streams will be numbered (in analogy with p. 76 of ref. [1]), so that

$$p_1 a_1 = b_1 c_1 \geq p_2 a_2 = b_2 c_2 \geq \ldots \geq p_N a_N = b_N c_N > 0. \tag{10}$$

Observe first that at a local minimum,

$$c_i x_i < c_j x_j \qquad \text{implies } x_j = 0 \qquad \text{or } x_i = b_i,$$

for otherwise we could further decrease the objective function (6) without violating the constraints (7) and (8), by decreasing x_j and increasing x_i by the same sufficiently small positive quantity. Since $x_j = 0$ in this situation would lead to a contradiction of the condition $x_j \geq 0$, we in fact have

$$c_i x_i < c_j x_j \qquad \text{implies } x_i = b_i. \tag{11}$$

In analogy with eq (4a) of reference [1], let r be the smallest index for which $x_r = b_r$ in the locally optimal solution under consideration. (If $x_i < b_i$ for $i = 1, 2, \ldots, N$, then take $r = N+1$.) We next show that

$$x_i = b_i \qquad \text{if} \qquad i \geq r, \tag{12}$$

i.e., that streams 1 through $r-1$ are precisely those served longer than needed to eliminate their queues. That $x_i < b_i$ for $i < r$, follows from the definition of r. To rule out the possibility that $x_i < b_i$ for some $i > r$, note that $b_i c_i \leq b_r c_r$, so that

$$c_i x_i < b_i c_i \leq b_r c_r = c_r x_r,$$

which by (11) implies $x_i = b_i$, a contradiction.

In particular, the solution is fully determined (each $x_i = b_i$) if $r = 1$, which by (12) and (9) can occur iff $fB = \sum_1^N b_i$. Thus in what follows we temporarily assume $B < \sum_1^N b_i$, so that $r > 1$.

Next, (11) and (12) imply the existence of some $K > 0$ such that

$$c_i x_i = K \qquad \text{for all } i < r,$$

or equivalently

$$x_i = q_i K \qquad \text{for all } i < r. \tag{13}$$

It follows from (8), (12), and (13) that

$$B = K \sum_1^{r-1} q_i + \sum_r^N b_i,$$

implying

$$K = \frac{(B - \sum_r^N b_i)}{\sum_1^{r-1} q_i}. \tag{14}$$

From (13) and the fact that $x_{r-1} < b_{r-1}$, we have

$$K < b_{r-1} c_{r-1} = p_{r-1} a_{r-1}. \tag{15}$$

If $r \leq N$, then it follows from $x_{r-1} < b_{r-1}$ and (11) . . . with $i = r-1$ and $j = r$ . . . that

$$K = c_{r-1} K q_{r-1} = c_{r-1} x_{r-1} \geq c_r x_r = c_r b_r (r \leq N). \tag{16}$$

We next dispose of the case $r = N+1$. By (8) and (13), if $r = N+1$ then

$$x_i = \frac{q_i B}{\sum_1^N q_j} \qquad \text{(all } i\text{)},$$

$$K = \frac{B}{\sum_1^N q_i}. \tag{17}$$

Using (15), we see from (18) that $r = N+1$ implies

$$B < b_N c_N \sum_1^N q_i. \tag{19}$$

Conversely if $r < N+1$, then (14) and (16) would both hold, yielding

$$B=K\sum_{1}^{r-1}q_i+\sum_{r}^{N}b_i\geq b_rc_r\sum_{1}^{r-1}q_i+\sum_{r}^{N}(b_ic_i)q_i\geq b_Nc_N\sum_{1}^{N}q_i,$$

contradicting (19). So (19) is a necessary and sufficient condition for $r=N+1$.

Suppose now that $1<r<N+1$. Using (14), (15), and (16), we have

$$b_rc_r\sum_{1}^{r-1}q_i+\sum_{r}^{N}b_i\leq B<b_{r-1}c_{r-1}\sum_{1}^{r-1}q_i+\sum_{r}^{N}b_i$$

as the test for determining r. With r known ($1<r<N+1$), the optimal solution is given by (12), (13), and (14). Since $b_{r-1}c_{r-1}q_{r-1}=b_{r-1}$, the test can be rewritten

$$B_r\leq B<B_{r-1}, \tag{20}$$

in terms of the quantities

$$B_k=b_kc_k\sum_{1}^{k-1}q_i+\sum_{k}^{N}b_i. \tag{21}$$

With the convention $B_{N+1}=0$, the test remains valid when $r=N+1$, according to the discussion surrounding (19). And with the convention $B_0=\infty$, it remains valid for $r=1$ as well (necessarily with $B=B_1$). The test is satisfied for at *most* one value of r since $B_{N+1}<B_N$ and for $1<k<N+1$,

$$B_{k-1}-B_k=(b_{k-1}c_{k-1}-b_kc_k)\sum_{1}^{k-1}q_i\geq 0;$$

it is satisfied for *at least* one value or r since $B_{N+1}<B\leq B_1$.

We conclude this section by summarizing the solution process, in terms of the problem data (assuming the ordering (10)):

Step 1: Calculate the total dead time per cycle, D.

Step 2: Calculate $B=N-1+(D/T)$.

Step 3: Calculate the quantities $b_i=(s_i-a_i)/s_i$ and their sum B_1.

Step 4: If $B>B_1$, then stop; the problem is infeasible. If $B=B_1$, the optimal solution is $G_i=Ta_i/s_i$ for all i. If $B<B_1$, continue.

Step 5: Beginning with B_1 and with $Q_0=0$, calculate quantities Q_1, B_2, Q_2, B_3, etc. by the formulas

$$q_k=\frac{b_k}{p_ka_k}$$

$$Q_k=Q_{k-1}+q_k,$$

$$B_k=B_{k-1}-(p_{k-1}a_{k-1}-p_ka_k)Q_{k-1}.$$

Stop as soon as $B_k\leq P$ is attained, set $r=k$, and go to Step 6. If $B<B_N$ is encountered, the optimal solution is

$$G_i=T-(TB/Q_N)q_i$$

for all i.

Step 6: Calculate

$$K=\frac{\left(B-\sum_{r}^{N}b_i\right)}{Q_{r-1}}.$$

The optimal solution is given by

$$G_i = T - TKq_i \qquad (i < r),$$

$$G_i = Ta_i/s_i \qquad (i \geq r).$$

4. Solution of Problem 2

Recall the relation

$$D/T = B - N + 1 \tag{22}$$

between B and T, which yields

$$dB/dT = -D/T^2. \tag{23}$$

The decreasing sequence $\{B_k\}_1^N$ defined by (21) yields, through (22), an increasing sequence $\{T_k\}_1^N$ of break-points in "T-space." The feasibility condition $B \leq B_1$ is equivalent to $T \geq T_1$, and the interval $B_k \leq B < B_{k-1}$ on which $r = k$ corresponds to the interval $T_{k-1} < T \leq T_k$.

Let $W_{\min}(T)$ be the minimized value of $W(T)$, as determined in section 3. Then by (5), we have the expression

$$W^0_{\min}(T) = T^2 W_{\min}(T)/2$$

for the minimum delay per cycle, so that

$$V^0(T) = TW_{\min}(T) \tag{24}$$

is twice the minimized *time-averaged* delay per cycle. Thus our objective in PROBLEM 2 is to choose T, subject to $T \geq T_1$, so as to minimize $V^0(T)$.

First consider the behavior of $V^0(T)$ on the interval (T_N, ∞) corresponding to the range $B < B_N$. By (17) and (6),

$$W_{\min}(T) = (B/Q_N)^2 \sum_1^N c_i q_i^2,$$

so that (24) yields

$$V^0(T) = (B^2 T) \times (\text{pos. const.}). \tag{25}$$

Using (23), we have

$$(d/dT)(B^2T) = B^2 + 2BT(dB/dT) = B^2 - 2B(D/T)$$

$$= B[2(N-1) - B] > B[2(N-1)) - B_N],$$

and since

$$B_N \leq B_1 = \sum_1^N b_i \leq N < 2(N-1)$$

(assuming of course that $N > 1$), it follows that (T_N, ∞) is an interval on which $V^0(T)$ is increasing, hence *not* an interval on which the minimum of $V^0(T)$ can occur.

Next, consider the behavior of $V^0(T)$ on the interval (T_{r-1}, T_r). Minimizing $V^0(T)$ over this interval is equivalent to minimizing

$$V_r(T) = V^0(T) \sum_1^{r-1} q_i. \tag{26}$$

Using (12) and (13) to substitute the optimal solution to PROBLEM 1 into (6), we obtain

$$W_{\min}(T) = K^2 \sum_1^{r-1} c_i q_i^2 + \sum_r^N c_i b_i^2$$

$$= K^2 \sum_1^{r-1} q_i + \sum_r^N c_i b_i^2,$$

which by (14) can be rewritten

$$W_{\min}(T) = \left(B - \sum_r^N b_i\right)^2 \left(\sum_1^{r-1} q_i\right)^{-1} + \sum_r^N c_i b_i^2. \tag{27}$$

It follows from (24) and (26) that

$$V_r(T) = T\left(B - \sum_r^N b_i\right)^2 + T\left(\sum_r^N c_i b_i^2\right)\left(\sum_1^{r-1} q_i\right).$$

This formula, (22) and (23) yield

$$dV_r/dT = \left(B - \sum_r^N b_i\right)\left(2N - 2 - \sum_r^N b_i - B\right) + \left(\sum_r^N c_i b_i^2\right)\left(\sum_1^{r-1} q_i\right), \tag{28}$$

$$d^2V_r/dT^2 = 2D^2/T^3 > 0. \tag{29}$$

Suppose in particular that $r \geqslant 3$. It will be shown that

$$dV_r/dT \geqslant 0 \text{ (right derivative at } T = T_{r-1}), \tag{30}$$

which by (29) implies that (T_{r-1}, T_r) is an interval on which $V_r(T)$ and hence $V^\circ(T)$ is increasing, hence *not* an interval on which the minimum of $V^\circ(T)$ can occur.

By (21) and (28), the expression in (30) whose sign is to be determined is

$$\left(B_{r-1} - \sum_r^N b_i\right)\left(2N - 2 - \sum_r^N b_i - B_{r-1}\right) + \left(\sum_r^N c_i b_i^2\right)\left(\sum_1^{r-1} q_i\right)$$

$$= \left(\sum_1^{r-1} q_i\right)\left\{b_{r-1}c_{r-1}\left(2N - 2 - 2\sum_r^N b_i - b_{r-1}c_{r-1}\sum_1^{r-1} q_i\right) + \sum_r^N c_i b_i^2\right\}.$$

This has the same sign as

$$D_{r-1} = 2N - 2 - 2\sum_r^N b_i - b_{r-1}c_{r-1}\sum_1^{r-1} q_i + (b_{r-1}c_{r-1})^{-1}\sum_r^N c_i b_i^2.$$

Since

$$b_{r-1}c_{r-1}\sum_1^{r-1} q_i \leqslant \sum_1^{r-1} b_i c_i q_i = \sum_1^{r-1} b_i,$$

we have

$$D_{r-1} \geqslant 2N - 2 - 2\sum_r^N b_i - \sum_1^{r-1} b_i + (b_{r-1}c_{r-1})^{-1}\sum_r^N c_i b_i^2.$$

Because each $b_i < 1$, the two subtracted terms in the right-hand side total less than

$$2(N - r + 1) + (r - 1) = 2N - r + 1,$$

which is no greater than $2N - 2$ for $r \geqslant 3$. Thus $D_{r-1} > 0$ for $r \geqslant 3$, verifying (30).

We have shown that the minimum of $V^{\circ}(T)$ over (T_1, ∞) is given by its minimum over $[T_1, T_2]$. That is, as noted in reference [1], in an optimal solution one has $r=1$ or $r=2$, so that for all but at most one stream one has $G_i/T = a_i/s_i$, i.e., all "slack time" (if there is any) is concentrated in the period allotted to a single stream.

The minimum over $[T_1, T_2]$ is determined as follows. Using (27) with $r=1$ and $r=2$, it is readily verified that $W_{\min}(T)$ is right-hand continuous at T_1. Thus the problem is equivalent to that of minimizing $V_2(T)$ over $[T_1, T_2]$.

By (29), the minimum will occur at T_1 if (30) applies there, and by (28) this condition reads

$$b_1\left(2N-2-b_1-2\sum_2^N b_1\right)+q_1\sum_2^N c_i b_i^2 \geq 0,$$

or equivalently

$$p_1a_1(2N-2-2B_1+b_1)+\sum_2^N p_ia_ib_i \geq 0. \tag{31}$$

If (31) does *not* hold, then $dV_2/dT=0$ occurs at a unique value T^*, where $T^* > T_1$, and the optimum occurs at T^* or T_2 according as $T^* \leq T_2$ or $T^* > T_2$. Specifically, from (28) and (22) we have

$$\left(D/T^*+N-1-\sum_2^r b_i\right)\left(N-1-D/T^*-\sum_2^r b_i\right)+\left(\sum_2^N c_ib_i^2\right)q_1=0,$$

or equivalently

$$\left(N-1-\sum_2^N b_i\right)^2-(D/T^*)^2+\left(\sum_2^N c_ib_i^2\right)q_1=0,$$

yielding

$$T^*=D\left\{\left(N-1-\sum_2^N b_i\right)^2+q_1\sum_2^N c_ib_i^2\right\}^{-1/2}. \tag{32}$$

The solution process for PROBLEM 2 can be summarized as follows, assuming the ordering (10):

Step 1: Calculate the total dead time per cycle, D.

Step 2: Calculate the quantities $b_i=(s_i-a_i)/s_i$, their sum B_1, and the quantity

$$B_2=B_1-(b_1-p_2a_2b_1/p_1a_1).$$

If $B_1 \leq N-1$, stop; the system is infeasible.

Step 3: If (31) holds, set $T=D/[B_1-(N-1)]$ and $G_i=Ta_i/s_i$ for all i.

Step 4: Otherwise, calculate $T_2=D/[B_2-(N-1)]$ and

$$T^*=D\left\{N-1-B_1+b_1)^2+(b_1/p_1a_1)\sum_2^N p_ia_ib_i\right\}^{-1/2}.$$

If $T^* > T_2$, set $T=T_2$ and

$$G_1=T_2(1+B_1-B_2-b_1).$$

If $T^* \leq T_2$, set $T=T^*$ and

$$G_1=T^*[B_1-b_1-(N-2)]-D.$$

In both cases, set $G_i=Ta_i/s_i$ for $i>1$.

5. References

[1] Rangarajan, R., and Oliver, R. M., Allocations of servicing periods that minimize average delay for N time-shared traffic streams, Transp. Sci. **1**, 74–80 (1967).

[2] Horn, W. A., Allocating service periods to minimize delay time, J. Res. Nat. Bur. Stand. (U.S.), **72B** (Math. Sci.), No. 3, 215–227 (1968).

(Paper 76B3&4–365)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Vol. 76B, Nos. 3 and 4, July–December 1972

Normal Matrices With Entries from an Arbitrary Field of Characteristic $\neq 2$

Martin H. Pearl and Alan I. Penn*

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

(July 11, 1972)

Let F be an arbitrary field of characteristic $\neq 2$ and let conjugation in F be defined by an arbitrary involutory automorphism (possibly the identity mapping). A matrix with entries from F is normal if it commutes with its conjugate transpose. Several conditions which are equivalent to normality when F is the complex field are properly nested when F is an arbitrary field. Of these conditions, normality is the weakest and unitary diagonalizability is the strongest.

When the underlying field is closed the unitarily diagonalizable matrices are those which possess a spectral representation with Hermitian idempotents. These matrices may be characterized in terms of their EP properties.

A matrix is indecomposable if it has a single elementary divisor of the form $(x-s)^i$. For $i>1$, **normal indecomposable matrices do not exist when F is the complex field. However there exist fields** for which normal indecomposable matrices of all finite orders exist. A matrix is a normal indecomposable matrix if and only if it can be expressed as $r(B)$ where $r(x)$ is a polynomial such that $r'(0)\neq 0$ and B is a matrix having the single elementary divisor x^i and satisfying the equation $B^*=sB$ for some scalar s.

When the involutory automorphism of F is the identity mapping indecomposable normal matrices of even order exist if and only if the vector space F^n is hyperbolic, and in this case the matrices are symmetric. Moreover, indecomposable normal matrices of odd order exist if and only if F^n is the orthogonal sum of a hyperbolic space and a one dimensional space, and in this case both symmetric and nonsymmetric indecomposable normal matrices exist.

Key words: Field; matrices; normality; hyperbolic space.

1. Introduction

Since Toeplitz [13][1] introduced normal matrices in 1918, there has been considerable effort to study their structure and properties. Several conditions equivalent to normality over the complex field have been found. In addition a number of generalizations of normality over the complex field have been proposed and investigated. Recently there has been an increased interest in the study of normal matrices over arbitrary base fields.

According to a classical theorem the following conditions are equivalent over the complex field [2]:

(1) A is normal, that is $A^*A=AA^*$

(2) A^* can be expressed as a polynomial in A with scalar coefficients.

(3) A has a spectral representation with Hermitian idempotents, that is, if $s_1, s_2, \ldots, s_k$ are the distinct characteristic roots of A then there exist matrices $E_1, E_2, \ldots, E_k$ such that:

(a) $I=\sum_i E_i$

(b) $E_iE_j=0$ for $i\neq j$

AMS Subject Classification: Primary 1535, Secondary 1545.

*Present address: University of Maryland, College Park, Md. 20740.

The work in this paper forms a portion of the second author's doctoral dissertation presented to the University of Maryland at College Park in 1971.

[1] Figures in brackets indicate the literature references at the end of this paper.

(c) $E_i^* = E_i$

(d) $A = \sum_i s_i E_i$

(4) A is unitarily similar to a diagonal matrix.

Additional conditions which are equivalent to normality over the complex field are given in [12, p. 222], [2, 4].

Generalizations to normality over the complex field have been proposed by Williamson [14, 15] and Schwerdtfeger [12]. Williamson observed the equivalence of conditions (1) and (2) above and generalized the concept of normality as follows: a matrix is *normal with respect to the Hermitian matrix* H if there exists a polynomial with scalar coefficients, $f(x)$, such that $AH = Hf(A^*)$. Schwerdtfeger generalized normality by introducing the concept of an EPr matrix: a matrix A is EPr if A has rank r and A and A^* have the same null space. That this is a generalization of normality over the complex field is most easily seen by noting that A is EPr if and only if A is unitarily similar to the direct sum of a nonsingular $r \times r$ matrix and a zero matrix [6].

In [9, 10] the concept of a normal matrix was extended to an arbitrary field with an involutory automorphism. Over an arbitrary field the notions of EPr and normality are independent. Moreover, both normality and the EPr property are necessary for A^* to be expressable as a unitary multiple of A ($A^* = UA = AU$ where $UU^* = I$), a condition which is weaker than unitary diagonalizability. Katz and Pearl [6] extended the results on normal EPr matrices over arbitrary fields, and found necessary and sufficient conditions for such matrices to satisfy the condition that A^* be a unitary multiple of A.

In this paper normal matrices will be studied for two extreme but important cases. In section 4, a characterization of normal matrices which are unitarily similar to a diagonal matrix is given for base fields which are closed and of characteristic $\neq 2$. Conjugation is defined by an arbitrary involutory automorphism. In section 5, we study normal matrices which have a single elementary divisor of the form $(x-s)^i$. Again the base fields are of characteristic $\neq 2$, and conjugation is defined by an arbitrary involutory automorphism. It is shown in Theorem 10 that matrices of this second form are normal if and only if they satisfy condition (2). While condition (4) precludes the existence of normal matrices of this second form for $i > 1$ over the complex field, there exist base fields for which such normal matrices exist for all orders. It is demonstrated in section 6 that one example of such a field is the complex numbers with conjugation defined by the identity automorphism. In section 6 more extensive results for the second case are obtained when conjugation is restricted to be the identity automorphism.

2. The Structure of Normal Matrices

Conditions (2), (3), and (4) define several key features of the structure of normal matrices over the complex field. However, when arbitrary base fields with associated involutory automorphisms are considered each of these conditions is stronger than that of normality, and consequently can not be used to define a structure on the full set of normal matrices. The proof of the following theorem is known but is included for completeness. (The conjugate transpose of the matrix M is denoted by M^*; the conjugate, or automorphic image, of the scalar s is denoted by $\bar{s}$).

THEOREM 1: *The following implications hold for an arbitrary base field:*

$$(4) \Rightarrow (3) \Rightarrow (2) \Rightarrow (1).$$

PROOF: (4) $\Rightarrow$ (3): There exists a diagonal matrix D and a unitary matrix U such that $A = UDU^*$. Consequently we can write A in the form

$$A = U \operatorname{diag}(s_1 I_{m_1}, s_2 I_{m_2}, \ldots, s_k I_{m_k}) U^*$$

where I_{m_j} is the unit matrix of order m_j. Let D_i denote the diagonal matrix obtained from diag $(s_1 I_{m_1}, s_2 I_{m_2}, \ldots, s_k I_{m_k})$ by replacing s_i by 1 and s_j by 0 for $j \neq i$. If we define $E_i = UD_iU^*$ then condition (3) is readily verified.

(3) $\Rightarrow$ (2): It is readily verified that $p(A) = \Sigma p(s_i)E_i$ for any polynomial $p(x)$. If we define $p_i(x) = (x-s_1)(x-s_2)\cdots(x-s_k)/(x-s_i)$ then we obtain $p_i(A) = p_i(s_i)E_i$. Consequently, E_i is expressible as a polynomial in A, and since $A^* = \Sigma \bar{s}_i E_i$ it follows that A^* is expressible as a polynomial in A.

(2) $\Rightarrow$ (1): Trivial. *Q.E.D.*

Counterexamples for the converse implications are given in Examples 1, 2, and 3.

EXAMPLE 1: Matrix satisfying (1) but not (2). Let the base field be $GF(7)$ with the identity automorphism. (Example taken from reference [9].) Let

$$B = \begin{pmatrix} 1 & 3 & 0 & 2 \\ 2 & -1 & 0 & 4 \\ 3 & 2 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

It is readily seen that $B^2 = 0$ and hence the only polynomials in B are of the form $pI + qB$. By considering the elements in the (1, 1) and (4, 4) positions of B, it is seen that if (2) holds, then $p = 0$ and $q = 1$ which in turn implies that B is symmetric. Since B is not symmetric (2) cannot hold. Condition (1) is verified by computation.

EXAMPLE 2: Matrix satisfying (2) but not (3). Let the base field be $GF(3)$ with the identity automorphism.
Let

$$C = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

The matrix C is symmetric and consequently satisfies condition (2). However C is nilpotent and consequently not diagonalizable. It remains to observe that any matrix with a spectral representation (with or without Hermitian idempotents) is diagonalizable [2]. Hence C cannot satisfy condition (3).

EXAMPLE 3: Matrix satisfying (3) but not (4). Let the base field be $GF(5)$ with the identity automorphism.
Let

$$D = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

It is readily verified that the matrices:

$$\begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}, \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix}$$

satisfy the conditions for the spectral representation and are symmetric. The matrix D has eigenvalues 3 with an associated eigenspace generated by $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and 1 with an associated eigenspace generated by $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Since neither eigenspace has a vector with unit length, there cannot exist a unitary transform which carries D into a diagonal matrix.

It should be noted that the reason the matrix D of Example 3 is not unitarily diagonalizable is because the base field is not large enough. If, in fact, the base field were algebraically closed this type of situation could not occur.

THEOREM 2: *Let* F *be an algebraically closed field of characteristic* $\neq 2$ *and* A *a matrix over* F. *Then condition* (4) *and condition* (3) *are equivalent.*

PROOF: That (4) implies (3) was shown above. Assume that A has a spectral decomposition with Hermitian idempotents. Then A must be diagonalizable: $A=BDB^{-1}$. Assume that $D=\text{diag}(s_1I_{m_1}, s_2I_{m_2}, \ldots, s_kI_{m_k})$ and define, as in Theorem 1, D_i to be the matrix obtained by using 1 in place of s_i and 0 in place of s_j for $j \neq i$. If we let $E_i=BD_iB^{-1}$ we obtain a spectral representation. Since the principal idempotents are uniquely determined, [2] the E_i must be Hermitian. Hence:

$$E_i=BD_iB^{-1}=(B^{-1})^*D_iB^*.$$

Consequently, the matrix B^*B must commute with D_i for each i. Thus, B^*B has a block diagonal form with the ith block corresponding to, and having the same dimension as, D_i. The m_i vectors of B which correspond to the ith block of B^*B define the eigenspace associated with s_i. Since there exists an orthogonal basis to this eigenspace, [8] we can assume that the m_i vectors of B in question are orthogonal. There is no loss of generality in assuming that B^*B is diagonal by applying the above arguments to each eigenspace. The matrix B is nonsingular, so that each of the diagonal elements of B^*B must be nonzero. Since the base field is assumed to be large enough to contain the square roots of each diagonal element of B^*B, the eigenvectors can be assumed to be normalized so that B is unitary. *Q.E.D.*

In order to gain insight into the structure of normal matrices over an arbitrary field two problems are treated. The first is studied in section 4, namely, to characterize the normal matrices which are unitarily diagonalizable. Since this is a stronger condition than (1), (2), or (3) we already have several necessary conditions; if we assume the base field is closed and of characteristic $\neq 2$, then condition (3) also becomes sufficient. It thus appears that the set of unitarily diagonalizable matrices over a closed field of characteristic $\neq 2$ is sufficiently small to permit meaningful analysis and sufficiently large to include as a special case the normal matrices over the complex field. The concept of EPr has proven to be a useful tool in the analysis of normality, and will be a key element in the study.

At the other extreme, normal matrices need not be diagonalizable, no less unitarily diagonalizable (see Example 2). The second problem, treated in section 5, is a characterization of the most fundamental matrices of this extremity, namely, those matrices which are normal and similar to a single Jordan block. It is shown in section 5 that when conjugation is defined by the identity such matrices have a well-defined structure dependent on the parity of the order: those matrices which are of even order must be symmetric; those of odd order may be nonsymmetric but must adhere to a specified form. In addition the existence of such matrices yields interesting conclusions about the geometry of the underlying vector space.

3. EP Matrices

A matrix A which satisfies the condition:

$$Ax=0 \quad \text{if and only if} \quad A^*x=0$$

is called an EP matrix. If A has rank r, then A is an EPr matrix.

It has been shown that the EP condition is necessary for unitary diagonalizability [12]. However, every nonsingular matrix is EP. Consequently, this condition is not sufficient. In this section the known results on EP matrices are reviewed and a more restrictive class of matrices, the universal EP matrices, will be defined. Finally, the known characterization of normal EP matrices will be reviewed and a characterization of normal universal EP matrices will be given.

THEOREM 3 [6, 7]: *The following statements are equivalent over an arbitrary field:*

(1) A *is an* EP *matrix.*

(2) A *is congruent to the direct sum of a nonsingular matrix* D *and a zero matrix.*

(3) *There exists a matrix* N *such that* $A^* = NA$.

(4) *There exists a nonsingular matrix* N *such that* $A^* = NA$.

(5) A *can be represented as:*

$$A = P\begin{pmatrix} D & DX^* \\ XD & XDX^* \end{pmatrix} P^* = P\begin{pmatrix} I & 0 \\ X & I \end{pmatrix}\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} I & X^* \\ 0 & I \end{pmatrix} P^*$$

where P *is a permutation matrix and* D *is a square nonsingular matrix.*

(6) A *and* A^* *have the same range space.*

From the definition of *EP* and condition (2) of Theorem 3 it follows that:

THEOREM 4: *Let* F *and* K *be fields,* $F \subseteq K$, *and let* A *be a matrix with entries from* F. *Then* A *is* EP *with respect to* F *if and only if* A *is* EP *with respect to* K.

When normality and *EP* are combined the following is obtained:

THEOREM 5 [10]: *A necessary and sufficient condition that* A *be a normal* EP *matrix is that there exists a nonsingular matrix* M *such that:*

$$A^* = MA = AM.$$

It is convenient to use the following variant of *EP*:

DEFINITION 1: Let A be a matrix with entries from a field F which contains all the characteristic roots of A. If every polynomial in A with coefficients from F is an *EP* matrix, then A is a *universal* EP *matrix.*

Example 4 demonstrates that universal *EP* is more restrictive than *EP*. However, from condition (2) of Theorem 3 it is readily seen that universal *EP* is still necessary for unitary diagonalizability.

EXAMPLE 4: Matrix *EP* but not universal *EP*. Let the base field be the algebraic closure of $GF(5)$ with identity automorphism.

$$E = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

The matrix E is nonsingular, hence *EP*. However, $E+I$ is not *EP*.

The following theorem characterizes the normal universal *EP* matrices:

THEOREM 6: *Let* A *be a normal matrix with entries from a field* F *which contains all of the characteristic roots* $s_1, s_2, \ldots, s_k$ *of* A. *If each of the polynomials*

$$A - s_i I \qquad i = 1, 2, \ldots, k$$

is an EP *matrix then* A *is a universal* EP *matrix.*

PROOF: Let

$$p(x) = p_m x^m + p_{m-1} x^{m-1} + \ldots + p_0$$

be an irreducible polynomial and set

$$\bar{p}(x) = \bar{p}_m x^m + \bar{p}_{m-1} x^{m-1} + \ldots + \bar{p}_0.$$

Then

$$[p(A)]^* = \bar{p}(A^*).$$

479-206 O - 73 - 3

CASE 1: Let $p(x)=c(x-s_i)$ for some i, $1 \leq i \leq k$, $c \in F$, $c \neq 0$. Then

$$p(A)=c(A-s_iI)$$

$$[p(A)]^*=\bar{c}(A^*-\bar{s}_iI).$$

By Theorem 5, there exists a nonsingular matrix M_i such that

$$A^*-\bar{s}_iI=M_i(A-s_iI)=(A-s_iI)M_i.$$

Clearly

$$[p(A)]^*=\bar{c}c^{-1}M_ip(A)=p(A)\bar{c}c^{-1}M_i.$$

Set $M_p=\bar{c}c^{-1}M_i$. Then M_p commutes with A and

$$[p(A)]^*=M_pp(A)=p(A)M_p.$$

CASE 2: Let $p(x)$ not be a multiple of any of the polynomials $x-s_i$, $i=1, 2, \ldots, k$. Then $p(A)$ is nonsingular. Set

$$M_p=[p(A)]^*[p(A)]^{-1}.$$

Since A is normal

$$[p(A)]^*=M_pp(A)=p(A)M_p$$

and M_p commutes with A.

Now let $f(x)$ be any polynomial and express $f(x)$ as a product of irreducible polynomials

$$f(x)=p_1(x)^{e_1}p_2(x)^{e_2} \ldots p_r(x)^{e_r}.$$

Then

$$[f(A)]^*=[p_1(A)^{e_1}]^*[p_2(A)^{e_2}]^* \ldots [p_r(A)^{e_r}]^*.$$

Defining M_{p_i} as in Case 1 we obtain

$$[f(A)]^*=M_ff(A)$$

where

$$M_f=M_{p_1}^{e_1}M_{p_2}^{e_2} \ldots M_{p_r}^{e_r}.$$

Clearly M_f is nonsingular and $f(A)$ is an EP matrix. *Q.E.D.*

4. Unitarily Diagonalizable Matrices

In this section criteria are given which, when combined with the universal EP property, form necessary and sufficient conditions for unitary diagonalizability over an arbitrary algebraically closed field of characteristic $\neq 2$.

Let A be a square matrix of order n with entries from F, and let $m(x)$ be the minimal polynomial of A. Express $m(x)$ as

$$m(x)=p_1(x)^{e_1}p_2(x)^{e_2} \ldots p_s(x)^{e_s} \tag{4-1}$$

where $p_1(x)$, $p_2(x)$, $\ldots$, $p_s(x)$ are distinct irreducible polynomials. Let us denote the vector space of all n dimensional column vectors with entries in F by F^n and let $N(B)$ denote the column null space of the matrix B, that is,

$$N(B)=\{x \mid x \in F^n, Bx=0\}.$$

According to the Primary Decomposition Theorem [5]

$$F^n = N[p_1(A)^{e_1}] \oplus N[p_2(A)^{e_2}] \oplus \ldots \oplus N[p_s(A)^{e_s}].$$

For vectors x and y belonging to F^n, we say that x and y are orthogonal if $x^*y = 0$. Two subspaces U and V of F^n are orthogonal if $x^*y = 0$ for all $x \in U$, $y \in V$. When U and V are orthogonal and have a zero intersection, we write

$$U \perp V.$$

LEMMA 1: *Let* A *be a normal matrix, let* r(x) *and* q(x) *be relatively prime polynomials and let* q(A) *be an* EP *matrix. Then*

$$\mathrm{N[r(A)]} \perp \mathrm{N[q(A)]}.$$

PROOF: Since $r(x)$ and $q(x)$ are relatively prime, it is clear that

$$N[r(A)] \cap N[q(A)] = 0. \tag{4-2}$$

It remains to show that $N[r(A)]$ and $N[q(A)]$ are orthogonal. Let $x \in N[r(A)]$, $y \in N[q(A)]$. Then

$$0 = x^*[q(A)y] = [x^*q(A)]y = [q(A)^*x]^*y.$$

Thus $q(A)^*N[r(A)]$ is orthogonal to $N[q(A)]$. Since A is a normal matrix, it follows that $N[r(A)]$ is A^*-invariant and hence

$$q(A)^*N[r(A)] = \bar{q}(A^*)N[r(A)] \subset N[r(A)]. \tag{4-3}$$

Since $q(A)$ is an *EP* matrix there exists a nonsingular matrix M such that

$$q(A)^* = Mq(A).$$

Thus,

$$q(A)^*N[r(A)] = Mq(A)N[r(A)]. \tag{4-4}$$

However, because $N[r(A)]$ is A-invariant, we have

$$q(A)N[r(A)] \subset N[r(A)]. \tag{4-5}$$

Also, it follows from (4-2) that

$$\dim\{q(A)N[r(A)]\} = \dim\{N[r(A)]\}. \tag{4-6}$$

Combining (4-5) and (4-6) we have

$$q(A)N[r(A)] = N[r(A)]. \tag{4-7}$$

Substituting (4-7) into (4-4) yields

$$q(A)^*N[r(A)] = MN[r(A)]. \tag{4-8}$$

Since M is nonsingular it follows that

$$\dim\{MN[r(A)]\} = \dim\{N[r(A)]\}. \tag{4-9}$$

Finally, combining (4–3), (4–8) and (4–9) we have

$$q(A)^*N[r(A)]=N[r(A)].$$

and hence $N[r(A)]$ is orthogonal to $N[q(A)]$. *Q.E.D.*

An immediate consequence of Lemma 1 is:

THEOREM 7: *Let* A *be a normal, universal* EP *matrix whose minimal polynomial is given by (4–1). Then*

$$F^n=N[p_1(A)^{e_1}]\perp N[p_2(A)^{e_2}]\perp\ldots\perp N[p_s(A)^{e_s}]. \tag{4–10}$$

EXAMPLE 5: Normal *EP* matrix for which Theorem 7 is false. The matrix E of Example 4 is readily verified to be normal *EP* with minimal polynomial $m(x)=(x-3)(x-4)$. The null spaces of the two irreducible factors are spanned by the isotropic vectors $\begin{pmatrix}1\\2\end{pmatrix}$ and $\begin{pmatrix}1\\3\end{pmatrix}$ respectively. Clearly, these null spaces are not orthogonal.

Since F^n is nonsingular and (4–10) is an orthogonal decomposition of F^n, each of the subspaces $N[p_i(A)^{e_i}]$, $i=1, 2, \ldots, s$ is also nonsingular [1]. Hence, if F is a field of characteristic $\neq 2$ then each $N[p_i(A)^{e_i}]$ has a basis of mutually orthogonal nonisotropic vectors. If we form the matrix U by taking for its columns the vectors of the bases for the subspaces $N[p_i(A)^{e_i}]$, $i=1, 2, \ldots, s$, then $UU^*=U^*U$ is a nonsingular diagonal matrix and

$$U^{-1}AU=B_1\oplus B_2\oplus\ldots\oplus B_s$$

where B_i is a square matrix whose order is equal to $\dim N[p_i(A)^{e_i}]$ and whose minimal polynomial is $p_i(x)^{e_i}$. In order that U be a unitary matrix it is necessary to be able to normalize each of the vectors in the bases of the subspaces $N[p_i(A)^{e_i}]$. If A is diagonalizable then each B_i has unit dimension, and A is unitarily diagonalizable.

THEOREM 8: Let F be an algebraically closed field of characteristic $\neq 2$ and let A be a square matrix with entries from F. Then a set of necessary and sufficient conditions that A be unitarily similar to a diagonal matrix is:

(1) A is similar to a diagonal matrix.

(2) A is normal.

(3) A is universal *EP*.

PROOF: The sufficiency follows from the above arguments. The necessity of (1) and (2) is clear. Since a diagonal matrix is an *EP* matrix, and since the *EP* property is invariant under congruence, it follows that every polynomial in A is an *EP* matrix. *Q.E.D.*

For a matrix A with entries from the complex field, Schwerdtfeger [12] has called the matrix A^*A the *Gram matrix* of A, and has shown that the Gram matrix of A always has the same rank as A. When the entries of A are from an arbitrary field, it is possible for A to have a greater rank than its Gram matrix (see Example 2).

We are now able to characterize the unitarily-diagonalizable matrices over suitable base fields without hypothesizing diagonalizability.

THEOREM 9: *Let* F *be an algebraically closed field of characteristic* $\neq 2$, *and let* A *be a square matrix with entries from* F. *Then a set of necessary and sufficient conditions that* A *be unitarily similar to a diagonal matrix is:*

(1) A *is universal* EP

(2) *For every* $c\in F$, $A-cI$ *has the same rank as its Gram matrix.*

PROOF: The necessity of (1) was shown in Theorem 8; the necessity of (2) is clear.

Let A be a matrix satisfying (1) and (2). Each of the polynomials $p_i(x)$ of (4–1) is linear since F is algebraically closed. Let $s_1, s_2, \ldots, s_k$ be the characteristic roots of A. By Theorem 3 there exists, for each i, a nonsingular matrix M_i such that:

$$(A-s_iI)^*=M_i(A-s_iI).$$

Hence:

$$(A-s_iI)^*(A-s_iI)=M_i(A-s_iI)^2.$$

Consequently, by (2), the matrices $(A-s_iI)^2$ and $(A-s_iI)$ have the same rank. Thus s_i is a simple characteristic root of A, and each of the exponents e_i of (4–1) is one. Hence A is diagonalizable, and F^n has a basis consisting of characteristic vectors of A. However, if x is a characteristic vector of A corresponding to any characteristic root s then:

$$(A^*-\bar{s}I)x=(A-sI)^*x=0.$$

Thus:

$$AA^*x=\bar{s}Ax=\bar{s}sx=A^*(sx)=A^*Ax$$

from which we have:

$$(AA^*-A^*A)x=0$$

and consequently A is normal. *Q.E.D.*

5. Normal Matrices Similar to a Single Jordan Block

In this section normal matrices having a single elementary divisor of the form $(x-s)^i$ are characterized and studied. The base field, F, is of characteristic $\neq 2$, and has an arbitrary involutory automorphism defining conjugation.

DEFINITION 2: A matrix, A, is *indecomposable* if A has a single elementary divisor of the form $(x-s)^i$. That is, if A is similar to a single Jordan block of the following form:

$$K=\begin{pmatrix} s & 1 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 & 0 \\ 0 & s & 1 & 0 & \cdot & \cdot & 0 & 0 & 0 & 0 \\ 0 & 0 & s & 1 & \cdot & \cdot & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & s & \cdot & \cdot & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdot & \cdot & s & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdot & \cdot & 0 & s & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & s & 1 \\ 0 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 & s \end{pmatrix} \tag{5–1}$$

Note that in the above definition of indecomposability the matrix A not only has a single elementary divisor over the base field but also has a single elementary divisor over all extension fields. This is, therefore, a more restrictive definition than is sometimes found where indecomposability assumes only a single elementary divisor over the base field [11].

EXAMPLE 6: Normal indecomposable matrix. Let the base field be $GF(5)$ with indentity automorphism. Let

$$F=\begin{pmatrix} 0 & 3 & 4 & 2 & 1 \\ 2 & 1 & 3 & 4 & 2 \\ 1 & 3 & 4 & 0 & 0 \\ 2 & 1 & 0 & 1 & 3 \\ 1 & 3 & 0 & 3 & 4 \end{pmatrix}=\begin{pmatrix} 0 & 0 & 1 & 3 & 4 \\ 0 & 2 & 3 & 4 & 4 \\ 0 & 1 & 4 & 1 & 0 \\ 2 & 0 & 0 & 0 & 3 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 & 3 & 4 \\ 0 & 2 & 3 & 4 & 4 \\ 0 & 1 & 4 & 1 & 0 \\ 2 & 0 & 0 & 0 & 3 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}^{-1}$$

The normality of F is readily verified.

THEOREM 10: *Let* A *be an indecomposable matrix over* F. *Then the following are equivalent:*

(1) A *is normal.*

(2) *There exists a polynomial with scalar coefficients such that* $A^* = p(A)$.

PROOF: That condition (2) implies (1) is clear. It remains to prove the converse. Let K be the upper Jordan block corresponding to A, and let H be a similarity transform which carries A to K:

$$H^{-1}AH = K.$$

The normality of A implies that

$$(HKH^{-1})^*(HKH^{-1}) = (HKH^{-1})(HKH^{-1})^*$$

which can be written as

$$[(H^*H)^{-1}K^*(H^*H)]K = K[(H^*H)^{-1}K^*(H^*H)].$$

Hence the matrix $(H^*H)^{-1}K^*(H^*H)$ commutes with the Jordan block matrix K and equivalently commutes with the matrix U having 1's on the first superdiagonal and 0's elsewhere. However, the only matrices which commute with U are upper triangular and constant on the diagonal and all superdiagonals [3]. Such a matrix has the following form:

$$\begin{pmatrix} a & b & c & d & . & . & . & y & z \\ 0 & a & b & c & . & . & . & . & y \\ 0 & 0 & a & b & c & . & . & . & . \\ 0 & 0 & 0 & a & b & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & a & b & c \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & b \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a \end{pmatrix} \tag{5-2}$$

But a matrix of the form (5–2) also has the form of a polynomial in the matrix K where:

$$a = p(\lambda) \qquad b = p'(\lambda) \qquad c = \frac{p''(\lambda)}{2!} \qquad \text{etc.}$$

Thus the matrix $(H^*H)^{-1}K^*(H^*H)$ is expressible as a polynomial in K:

$$(H^*H)^{-1}K^*(H^*H) = p(K). \tag{5-3}$$

From eq (5–3) we obtain:

$$A^* = p(A). \qquad \textit{Q.E.D.}$$

EXAMPLE 7: Polynomial transform for normal indecomposable matrix. The matrix F of Example 6 satisfies the following equality:

$$F^* = F^4 + F^3 + 2F^2 - F.$$

On the basis of Theorem 10 we can characterize the indecomposable normal matrix A by the polynomial transformation which carries A into A^*. To a certain extent matrices have long been

characterized by this transformation; symmetric and skew-symmetric matrices are those in which the polynomial is the identity and negative identity, respectively. The following generalization of symmetry and skew-symmetry will be useful in the ensuing discussion:

DEFINITION 3: A matrix A is *unit-symmetric* if $A = uA^*$ for some scalar $u \neq 0$.

If the matrix A is not the zero matrix then $A = uA^*$ implies $A = (u\bar{u})A$ and consequently the scalar u must have unit modulus.

If the automorphism is the identity then a unit-symmetric matrix is either symmetric or skew-symmetric since the only scalars with unit modulus are $+1$ and -1; if, however, the automorphism is not the identity then the scalar u can lie anywhere on the "unit circle."

While the indecomposable matrix A can have as its eigenvalue an arbitrary scalar in the field, it will be convenient to always assume that the eigenvalue is 0. The next lemma shows that any results obtained under this assumption can be readily translated to the case where the eigenvalue is arbitrary.

LEMMA 2: *Let the matrix* A *be indecomposable and let* a *be an arbitrary scalar in the field. Then* A *is normal if and only if* (A−aI) *is normal. Consequently,* A* *is expressible as a polynomial in* A *if and only if* (A−aI)* *is expressible as a polynomial in* (A−aI).

PROOF: The proof consists of verifying the equivalence of the two equations:

(1) $A^*A = AA^*$

(2) $(A - aI)^*(A - aI) = (A - aI)(A - aI)^*$.

The next lemma establishes a relatively weak but important set of necessary conditions on the polynomials which transform a normal indecomposable matrix with 0 eigenvalue into its conjugate transpose.

LEMMA 3: *Let* $A^* = p(A)$ *where* A *is an* $n \times n$ *normal indecomposable matrix with* 0 *eigenvalue. Then the polynomial* p(x) *has the following form:*

$$p(x) = a_{n-1}x^{n-1} + \ldots + a_1 x$$

where $a_1 \neq 0$.

PROOF: Since $A^n = 0$ it is not necessary to consider any term in the polynomial beyond x^{n-1}. Since the eigenvalue 0 is in the fixed field of the involutory automorphism, the matrix A is similar to A^*. Consider A^* in the following form:

$$A^* = p(A) = p(HKH^{-1}) = Hp(K)H^{-1} \tag{5-4}$$

where K is the upper Jordan block corresponding to A, and H is the similarity transform which carries A into K. Since A^* is similar to A, it follows from equation (5−4) that $p(K)$ is also similar to A. Hence, $p(K)$ is indecomposable with 0 eigenvalue. The matrix $p(K)$ has the upper triangular form shown in (5−2). Since the eigenvalue of $p(K)$ is zero it follows that the diagonal element is zero. But the diagonal element of $p(K)$ assumes the value $p(0)$. Then

$$p(0) = 0.$$

Since $p(K)$ is indecomposable it has rank $n-1$, and consequently the element defining the first superdiagonal is nonzero. But this superdiagonal is defined by $p'(0)$.

$$p'(0) \neq 0. \qquad \textit{Q.E.D.}$$

DEFINITION 4: A polynomial $p(x)$ which is of the form

$$p(x) = a_{n-1}x^{n-1} + \ldots + a_1 x$$

where $a_1 \neq 0$ is called an *nth order admissible polynomial*, or where the order is understood, simply an *admissible polynomial.*

It should be kept in mind that the term *n*th order refers to the dimension of the matrix A upon which the polynomial acts, and that an *n*th order admissible polynomial has degree at most $(n-1)$. The identity polynomial is an *n*th order admissible polynomial for all *n*. If $p(x)$ is admissible then $\bar{p}(x)$ is also admissible where $\bar{p}(x)$ is obtained from $p(x)$ by taking the conjugate of each coefficient.

It is possible to modify the usual concept of composition of polynomials in such a way that it is unchanged on the class of admissible polynomials, but it is better behaved as an algebraic structure. This is the basis of the next definition.

DEFINITION 5: Let $p(x)$ and $q(x)$ be the *n*th order admissible polynomials. Then the n *modular composition of* p(x) *and* q(x) is defined to be the polynomial obtained by first taking the ordinary composition of $p(x)$ and $q(x)$ as polynomials over the base field, and then dropping all terms of degree greater than $(n-1)$. The modular composition of $p(x)$ and $q(x)$ is denoted by $(p(x) \cdot q(x))$ or $(p \cdot q)(x)$.

It is clear from the definition that the modular composition of two *n*th order admissible polynomials is again an *n*th order admissible polynomial. Moreover, if $p(x)$ and $q(x)$ are *n*th order admissible polynomials, and if A is an $n \times n$ indecomposable matrix with 0 eigenvalue then:

$$p(q(A)) = (p \cdot q)(A).$$

THEOREM 11: *For any integer* n > 1, *the* n*th order admissible polynomials under the operation of* n *modular composition form a group.*

PROOF: It has already been noted that the admissible polynomials are closed under the modular composition. Associativity follows from the associativity of polynomials under ordinary composition. The polynomial x acts as the identity and is admissible. Let $p(x)$ and $q(x)$ be admissible polynomials:

$$p(x) = p_{n-1}x^{n-1} + \ldots + p_1 x \qquad p_1 \neq 0$$

$$q(x) = q_{n-1}x^{n-1} + \ldots + q_1 x \qquad q_1 \neq 0.$$

In order that $q(x)$ be the left inverse of $p(x)$ the following equation must hold:

$$q_{n-1}(p_{n-1}x^{n-1} + \ldots + p_1 x)^{n-1} + \ldots + q_1(p_{n-1}x^{n-1} + \ldots + p_1 x) \doteq x \qquad (5\text{–}5)$$

where the symbol $\doteq$ refers to a congruent equality in which all terms above $(n-1)$ are ignored. Equation (5–5) is equivalent to the system of $(n-1)$ equations:

$$q_1 p_1 = 1$$

$$q_i p_1^i + f_i\,(q_1, \ldots, q_{i-1}, p_1, \ldots, p_{i-1}) + q_1 p_i = 0 \qquad i = 2, \ldots, (n-1) \qquad (5\text{–}6)$$

where f_i is a well-defined function of q_j, p_j for $j < i$.

From the form of eq (5–6) it is clear that for a given polynomial $q(x)$, the polynomial $p(x)$ is uniquely determined by a simple recursive relationship. Conversely, given $p(x)$, then the polynomial $q(x)$ is uniquely determined. Thus the left and right inverses of a given admissible polynomial exist and are unique. Hence the set of admissible polynomials under modular composition form a group. *Q.E.D.*

EXAMPLE 8: Group of admissible polynomials. The following is the group table of the third order admissible polynomials over $GF(3)$.

	x	$2x$	x^2+x	x^2+2x	$2x^2+x$	$2x^2+2x$
x	x	$2x$	x^2+x	x^2+2x	$2x^2+x$	$2x^2+2x$
$2x$	$2x$	x	$2x^2+2x$	$2x^2+x$	x^2+2x	x^2+x
x^2+x	x^2+x	x^2+2x	$2x^2+x$	$2x^2+2x$	x	$2x$
x^2+2x	x^2+2x	x^2+x	$2x$	x	$2x^2+2x$	$2x^2+x$
$2x^2+x$	$2x^2+x$	$2x^2+2x$	x	$2x$	x^2+x	x^2+2x
$2x^2+2x$	$2x^2+2x$	$2x^2+x$	x^2+2x	x^2+x	$2x$	x

Since this group is nonabelian it is isomorphic to S_3. In general, the group of jth order admissible polynomials over $GF(n)$ has order $n^{j-2}(n-1)$.

The admissible polynomials, having been structured as a group, can now be partitioned according to the following equivalence relationship:

DEFINITION 6: The admissible polynomials $p(x)$ and $q(x)$ are *equivalent admissible polynomials* if there exists an admissible polynomial $r(x)$ such that

$$\bar{r}(x)^{-1} \cdot p(x) \cdot r(x) = q(x).$$

EXAMPLE 9: Equivalence classes of admissible polynomials. The following equivalence classes result from the third order admissible polynomials over $GF(3)$ (see Example 8).

(1) $\{x\}$
(2) $\{2x, x^2+2x, 2x^2+2x\}$
(3) $\{x^2+x, 2x^2+x\}$.

If a polynomial is to transform an indecomposable matrix with 0 eigenvalue into its conjugate transpose it must be admissible. However, this necessary condition can be strengthened considerably; that is, the group of admissible polynomials includes many polynomials which are incapable of transforming such a matrix into the conjugate transpose.

LEMMA 4: *Let the* n × n *matrix* A *be indecomposable with* 0 *eigenvalue and let* p(x) *be a polynomial such that* $A^* = p(A)$. *Then the polynomial* $(\bar{p} \cdot p)(x)$ *is the identity polynomial. Consequently* $p'(0)$ *has unit modulus.*

PROOF: Since $A^* = p(A)$ it follows that $A = \bar{p}(A^*)$. Hence:

$$A = \bar{p}(p(A)) = (\bar{p} \cdot p)(A).$$

This same polynomial identity holds for all matrices similar to A. In particular, it holds for the upper Jordan block matrix U consisting of 1's on the first superdiagonal and 0's elsewhere:

$$U = (\bar{p} \cdot p)(U). \tag{5-7}$$

From eq (5–7) it is seen that the first order term of $(\bar{p} \cdot p)$ has a unit coefficient and that all higher order terms are zero; hence, $(\bar{p} \cdot p)(x) = x$. Since $p(0) = 0$ (Lemma 3) it follows that:

$$(\bar{p} \cdot p)'(0) = \bar{p}'(0) \cdot p'(0).$$

Hence $p'(0)$ has unit modulus. *Q.E.D.*

THEOREM 12: *Let the* n × n *matrix* A *be indecomposable with* 0 *eigenvalue and let* p(x) *be a polynomial such that* $A^* = p(A)$. *Then* p(x) *is equivalent to the polynomial* $r(x) = x/p'(0)$ *where* $p'(0)$ *has unit modulus.*

PROOF: Define a polynomial $q(x)$ as follows:

$$q(x) = p(x) + p'(0)x.$$

Since the first order term of $q(x)$ has the coefficient $2p'(0) \neq 0$ it follows that $q(x)$ is admissible. From Lemma 4 it is known that $(\bar{p} \cdot p)(x) = x$. Hence:

$$(\bar{q}(x) - \overline{p'(0)}x) \cdot (p(x)) = x.$$

Applying the left distributive law which follows from this distributive property of polynomials, we obtain:

$$\bar{q}(x) \cdot p(x) = x + (\overline{p'(0)}x) \cdot (p(x)) = (\overline{p'(0)}x) \cdot (q(x)). \tag{5-8}$$

Hence:

$$p(x) = (\bar{q}(x))^{-1} \cdot (\overline{p'(0)}x) \cdot (q(x)). \qquad Q.E.D.$$

EXAMPLE 10: Equivalence of transforming polynomial to polynomial of form $x/p'(0)$. The transforming polynomial of Example 7, $p(x) = x^4 + x^3 + 2x^2 - x$, is equivalent to the polynomial $-x$:

$$(x^3 + x^2 + x)^{-1} \cdot p(x) \cdot (x^3 + x^2 + x) = -x.$$

EXAMPLE 11: Examples of normal indecomposable matrices corresponding to polynomials of form $x/p'(0)$. For base field $GF(3)$ the only polynomials equivalent to polynomials of the form $x/p'(0)$ are in equivalence classes (1) and (2) of Example 9. Matrices corresponding to these four polynomials are:

$$G_1 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 2 \\ 1 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 2 \\ 2 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 2 \\ 2 & 0 & 2 \end{pmatrix}^{-1} \qquad G_1^t = G_1$$

$$G_2 = \begin{pmatrix} 0 & 2 & 1 \\ 1 & 0 & 2 \\ 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix}^{-1} \qquad G_2^t = -G_2$$

$$G_3 = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 2 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}^{-1} \qquad G_3^t = G_3^2 + 2G_3$$

$$G_4 = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix}^{-1} \qquad G_4^t = 2G_4^2 + 2G_4.$$

In the above development, a normal indecomposable matrix with 0 eigenvalue is characterized by the polynomial which transforms the given matrix into the conjugate transpose. In the following development the attention is switched to the matrix itself.

LEMMA 5: *Let* A *be an* n × n *matrix and* p(x) *be an nth order admissible polynomial.*

(1) *If* A *is indecomposable with* 0 *eigenvalue then* p(A) *is also indecomposable with* 0 *eigenvalue.*

(2) *If* A *is normal then* p(A) *is also normal.*

PROOF: If A is indecomposable with 0 eigenvalue then A is similar to the matrix U having 1's on the first superdiagonal and 0's elsewhere. Hence $p(A)$ is similar to $p(U)$ where $p(U)$ has the form shown in (5–2). Since $p(0)=0$ it follows that the diagonal of $p(U)$ is zero and that the only eigenvalue of $p(U)$ is zero. Since $p'(0)\neq 0$ it follows that the first superdiagonal of $p(U)$ is nonzero and consequently, that $p(U)$ is indecomposable. Finally, since $p(A)$ is similar to $p(U)$ it follows that $p(A)$ is also indecomposable with 0 eigenvalue.

If A is normal then it is readily verified that $p(A)$ is normal for any polynomial $p(x)$. *Q.E.D.*

LEMMA 6: *Let the* n×n *matrix* A *be indecomposable normal with* 0 *eigenvalue. Then there exists an* n×n *unit-symmetric indecomposable matrix* B *with* 0 *eigenvalue and an admissible polynomial* r(x) *such that* A=r(B) *and* B=r^{-1}(A).

PROOF: Define the admissible polynomial $q(x)$ as in Theorem 12:

$$q(x)=p(x)+p'(0)x$$

where $p(x)$ is the polynomial which transforms A into A^*. Evaluating eq (5–8) at $x=A$ the following is obtained:

$$\bar{q}(p(A))=\overline{p'(0)}q(A). \tag{5–9}$$

Define the matrix B as follows:

$$B=q(A).$$

Since $q(x)$ is admissible the matrix B is indecomposable with 0 eigenvalue and normal by Lemma 5. The actual form of B^* can be evaluated as follows:

$$B^*=\bar{q}(A^*)=\bar{q}(p(A)). \tag{5–10}$$

From eqs (5–9) and (5–10) and the definition of B it follows that:

$$B^*=\overline{p'(0)}B.$$

Hence B is unit-symmetric.

Since $q(x)$ is admissible and the matrices B and A are indecomposable with 0 eigenvalue, it follows that $q^{-1}(x)$ can be applied to both sides of the equation defining B to obtain

$$A=q^{-1}(B).$$

The lemma is proved if we define the polynomial $r(x)=q^{-1}(x)$. *Q.E.D.*

EXAMPLE 12: Transformation of normal indecomposable matrix into unit-symmetric indecomposable matrix. The matrix F of Example 6 can be obtained from or transformed to the following skew-symmetric indecomposable matrix:

$$H=\begin{pmatrix}0&3&4&0&0\\2&0&0&0&0\\1&0&0&3&4\\0&0&2&0&0\\0&0&1&0&0\end{pmatrix}=\begin{pmatrix}0&0&1&0&1\\0&2&0&2&0\\0&1&0&0&0\\2&0&0&0&0\\1&0&0&0&1\end{pmatrix}\begin{pmatrix}0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\\0&0&0&0&0\end{pmatrix}\begin{pmatrix}0&0&1&0&1\\0&2&0&2&0\\0&1&0&0&0\\2&0&0&0&0\\1&0&0&0&1\end{pmatrix}^{-1}$$

The transformations are as follows:

$$F=H^3+H^2+H$$

$$H=F^3-F^2+F.$$

The results of Lemma 6 can be readily extended to obtain a complete characterization of normal indecomposable matrices with arbitrary eigenvalues.

THEOREM 13: *Let the matrix* A *be indecomposable. Then* A *is normal if and only if* A *can be represented in the form* $A = r(B)$ *where the matrix* B *is unit-symmetric indecomposable with* 0 *eigenvalue and* $r'(0) \neq 0$.

PROOF: Assume first that A is normal indecomposable with eigenvalue s. Then $(A - sI)$ is normal indecomposable with 0 eigenvalue. By Lemma 6 there exists a unit-symmetric indecomposable matrix B with 0 eigenvalue and an admissible polynomial $q(x)$ such that:

$$(A - sI) = q(B).$$

Hence:

$$A = q(B) + sI = r(B).$$

Since $q(x)$ is admissible it follows that $q'(0) \neq 0$. Hence $r'(0) \neq 0$. The converse follows from the fact that any polynomial in a normal matrix is again normal. *Q.E.D.*

On the basis of Theorem 13 we can readily obtain all the indecomposable normal matrices from the set of indecomposable unit-symmetric matrices with 0 eigenvalue.

If the involutory automorphism which defines conjugation is the identity, then the problem is greatly simplified. For this special case the only unit-symmetric matrices are those which are either symmetric or skew-symmetric. Section 6 of this paper treats this case in detail.

6. Indecomposable Normal Matrices With Entries From a Field in Which Conjugation Is Defined by the Identity Mapping

In this section the results of section 5 are extended for the special case in which the involutory automorphism defining conjugation is the identity mapping. To distinguish the present results from those obtained for an arbitrary involutory automorphism the transpose of a matrix will be denoted by A^t rather than by A^*.

The analog to Theorem 12 assumes the following form:

THEOREM 14: *Let the matrix* A *be indecomposable with* 0 *eigenvalue and let* p(x) *be a polynomial such that* $A^t = p(A)$. *Then* $p'(0)$ *is either* $+1$ *or* -1. *If* $p'(0) = -1$ *then* p(x) *is equivalent to the negative identity polynomial,* $-x$. *If* $p'(0) = +1$ *then* p(x) *is the identity polynomial,* x.

PROOF: From Lemma 4, $p'(0)$ has unit modulus. Since the automorphism is the identity the only scalars with unit modulus are $+1$ and -1.

If $p'(0) = -1$ then $p(x)$ is equivalent to the negative identity by Theorem 12. If $p'(0) = +1$ then $p(x)$ is equivalent to the identity by the same theorem; however, the only polynomial which is equivalent to the identity polynomial is the identity itself. *Q.E.D.*

The characterization of the normal indecomposable matrices in terms of unit-symmetric matrices assumes the following form:

THEOREM 15: *Let the matrix* A *be indecomposable. Then* A *is normal if and only if one of the following conditions is satisfied:*

(1) A *is symmetric.*

(2) *There exists a skew-symmetric indecomposable matrix* B *with* 0 *eigenvalue and a polynomial* r(x) *where* $r'(0) \neq 0$ *such that:*

$$A = r(B).$$

Conditions (1) and (2) cannot be simultaneously satisfied.

PROOF: Assume first that A is normal. By Theorem 13 there exists a unit-symmetric indecomposable matrix B with 0 eigenvalue and a polynomial $r(x)$, $r'(0) \neq 0$, such that:

$$A = r(B).$$

Since the automorphism is the identity, the matrix B is either symmetric or skew-symmetric. If B is symmetric then A is symmetric; if B is skew-symmetric then condition (2) is satisfied.

The converse follows from the fact that a polynomial in a normal matrix is again normal.

It remains to show that conditions (1) and (2) can not be simultaneously satisfied. Suppose the matrix A satisfies both conditions. If A has order n then $B^n=0$ and we can assume that $r(x)$ has degree less than or equal to $n-1$:

$$A=r_{n-1}B^{n-1}+\ \ .\ .\ .+r_2B^2+r_1B^1+r_0I. \tag{6-1}$$

Since A is symmetric and B skew-symmetric the following equation is obtained by taking the transpose of eq (6–1):

$$A=\pm r_{n-1}B^{n-1}\mp .\ .\ .+r_2B^2-r_1B^1+r_0I. \tag{6-2}$$

If n is an odd integer subtract eq (6–2) from (6–1); if n is an even integer add the two equations to obtain:

$$0=2\{r_{n-2}B^{n-2}+r_1B\ (r_0I)\}. \tag{6-3}$$

Since B is indecomposable with 0 eigenvalue its minimal polynomial is given by x^n. But eq (6–3) implies that B is annihilated by a polynomial of degree $n-2$. Contradiction. *Q.E.D.*

It should be noted that any skew-symmetric indecomposable matrix has zero eigenvalue when conjugation is defined by the identity. For if B is skew-symmetric indecomposable with eigenvalue s, then B^t also has eigenvalue s. But $B^t=-B$, and consequently B^t has eigenvalue $-s$. Hence $s=0$.

The polynomials which relate a normal indecomposable matrix with 0 eigenvalue to its transpose are completely defined by the following set: the identity polynomial plus the set of polynomials in the equivalence class of the negative identity. A more explicit representation of this set of polynomials is desirable but has not been obtained.

THEOREM 16: *Let* A *be normal indecomposable. If* A *is not symmetric then* A *is of odd order. Symmetric indecomposable matrices exist for both odd and even orders.*

PROOF: The existence of symmetric matrices of odd and even orders and nonsymmetric matrices of odd order is shown by Examples 13, 14, 15. If A is normal indecomposable but not symmetric then by Theorem 15 there exists a skew-symmetric indecomposable matrix with 0 eigenvalue, say B, of the same order. The rank of B is clearly one less than the order since it is similar to a matrix having all 0's except 1's on the first superdiagonal. It remains to observe that any skew-symmetric matrix has even rank [11]. *Q.E.D.*

EXAMPLE 13: Example of symmetric indecomposable matrix of even order. Base field $GF(5)$.

$$\begin{pmatrix}1&3\\3&4\end{pmatrix}=\begin{pmatrix}2&2\\1&0\end{pmatrix}\begin{pmatrix}0&1\\0&0\end{pmatrix}\begin{pmatrix}2&2\\1&0\end{pmatrix}^{-1}.$$

EXAMPLE 14: Example of symmetric indecomposable matrix of odd order. Base field $GF(5)$.

$$\begin{pmatrix}0&0&2\\0&0&1\\2&1&0\end{pmatrix}=\begin{pmatrix}2&0&0\\1&0&1\\0&1&0\end{pmatrix}\begin{pmatrix}0&1&0\\0&0&1\\0&0&0\end{pmatrix}\begin{pmatrix}2&0&0\\1&0&1\\0&1&0\end{pmatrix}^{-1}.$$

EXAMPLE 15: Example of skew-symmetric indecomposable matrix of odd order. Base field. $GF(5)$.

$$\begin{pmatrix}0&0&2\\0&0&1\\3&4&0\end{pmatrix}=\begin{pmatrix}2&0&0\\1&0&4\\0&1&0\end{pmatrix}\begin{pmatrix}0&1&0\\0&0&1\\0&0&0\end{pmatrix}\begin{pmatrix}2&0&0\\1&0&4\\0&1&0\end{pmatrix}^{-1}.$$

Theorem 16 indicates a fundamental difference between the symmetric and nonsymmetric normal indecomposable matrices. Since any polynomial of a symmetric matrix is clearly symmetric, polynomial transforms are not useful for studying the interrelationship between the symmetric and the nonsymmetric matrices. More complex tools are needed for any coexistence theorems for these two classes.

Theorem 16 also demonstrates a fundamental difference between normal indecomposable matrices of odd and even order. Since this parity effect is one which is often apparent in geometric structures, a study of the geometric implications of the underlying vector space is indicated. Several definitions and preliminary results are needed for the discussion of geometric properties.

DEFINITION 7: For a vector space V, the *radical* of V, rad V, is the subspace of all vectors of V which are orthogonal to V. If U^* denotes the subspace of V orthogonal to U then rad $U = U \cap U^*$. We say that V is *nonsingular* if rad $V = 0$. If rad $V = V$ then V is *isotropic*.

DEFINITION 8: A 2-dimensional nonsingular vector space which contains an isotropic vector is called a *hyperbolic plane*. An orthogonal sum of hyperbolic planes is a *hyperbolic space*.

A hyperbolic space is nonsingular and has even dimension. As in section 4, we will use the symbol $V \perp W$ to denote the orthogonal sum of the spaces V and W.

THEOREM 17 [1]: *Let* V *be nonsingular and* U *any subspace of* V. *Write* $\mathrm{U} = rad\ \mathrm{U} \perp \mathrm{W}$ *and let* $\mathrm{N}_1, \mathrm{N}_2, \ldots, \mathrm{N}_r$ *be a basis of rad* U. *Then we can find in* V *vectors* $\mathrm{M}_1, \mathrm{M}_2, \ldots, \mathrm{M}_r$ *such that each* $\mathrm{N}_1, \mathrm{M}_1$ *is a hyperbolic pair and such that the hyperbolic planes* $\mathrm{P}_1 = (\mathrm{N}_1, \mathrm{M}_1)$ *are mutually orthogonal and also orthogonal to* W. V *will therefore contain the nonsingular space:*

$$\overline{\mathrm{U}} = \mathrm{P}_1 \perp \mathrm{P}_2 \perp \ldots \perp \mathrm{P}_r \perp \mathrm{W}$$

which, in turn, contains U.

THEOREM 18: *If* n *is an even integer then the following statements are equivalent:*

(1) F^n *is a hyperbolic vector space.*

(2) *There exist* $\mathrm{n} \times \mathrm{n}$ *symmetric indecomposable matrices over* F.

If n *is an odd integer, the following statements are equivalent:*

(1′) F^n *contains an* $\mathrm{n}-1$ *dimensional hyperbolic subspace.*

(2′) *There exist* $\mathrm{n} \times \mathrm{n}$ *symmetric indecomposable matrices over* F.

(3′) *There exist* $\mathrm{n} \times \mathrm{n}$ *skew-symmetric indecomposable matrices over* F.

(4′) *There exist* $\mathrm{n} \times \mathrm{n}$ *normal indecomposable matrices over* F.

PROOF: We will show first that the existence of any normal indecomposable matrix implies (1) for n even, or (1′) for n odd. Let A be a normal indecomposable matrix. If A is symmetric with eigenvalue s then $A - sI$ is indecomposable symmetric with 0 eigenvalue. If A is not symmetric then there exists, by Theorem 15, an indecomposable skew-symmetric matrix B which must have 0 eigenvalue. In either case we are assured of the existence of an indecomposable normal matrix with 0 eigenvalue.

Let $p(x)$ denote the polynomial transformation that carries the matrix B into B^t. Since B is either symmetric or skew-symmetric, the polynomial $p(x)$ will be correspondingly either the identity or the negative identity.

If H is the similarity transformation which carries B into the single Jordan block K:

$$B = HKH^{-1}$$

then the following is readily obtained:

$$(HKH^{-1})^t = B^t = p(B) = Hp(K)H^{-1}.$$

Hence:

$$(H^tH)p(K) = K^t(H^tH).$$

The matrix (H^tH) is the Gram matrix of the matrix H. Let L denote this Gram matrix so that the above assumes the form:

$$Lp(K)=K^tL.$$

Note that since H is a similarity transform it must be nonsingular; consequently the Gram matrix L is also nonsingular.

Since $p(x)$ is either the identity or the negative identity, the matrix $p(K)$ consists of all 0's except $+1$ or -1 on the first superdiagonal. Hence the first column of $Lp(K)$ is identically 0, and the first column of K^tL consists of 0 in the first place and $L_{i-1,1}$ in the ith place, $i>1$. Thus:

$$L_{1,1}=L_{2,1}=\ .\ .\ .=L_{n-1,1}=0. \tag{6–4}$$

Since L is nonsingular, $L_{n,1}$ is nonzero.

Examining the second column of $Lp(K)$ it is found that eqs (6–4) imply that there are 0's in the first $n-1$ places. Comparing this to the second column of K^tL the following is obtained:

$$L_{1,2}=L_{2,2}=\ .\ .\ .=L_{n-2,2}=0.$$

Continuing through the n-1st column we obtain:

$$(H^tH)_{i,j}=L_{i,j}=0 \text{ whenever } i+j\leq n. \tag{6–5}$$

If we consider a basis of F^n defined by the columns of the similarity transform H, h_i, $i=1, 2, \ .\ .\ ., n$, then the length square of the basis vector h_i is given by the ith element along the diagonal of the Gram matrix L. If n is an even integer then eq (6–5) implies that the first $\frac{n}{2}$ vectors of this ordered basis are isotropic. If n is an odd integer then the first $\frac{(n-1)}{2}$ vectors of this ordered basis are isotropic.

If we define the subspace U to be generated by the first $\frac{n}{2}$ vectors in the basis h_i $\left(\frac{(n-1)}{2} \text{ if } n \text{ odd}\right)$, then the space F^n is seen to be hyperbolic (contain an $n-1$ dimensional hyperbolic subspace if n odd).

Next assume that n is even and F^n is a hyperbolic space. Then there exists $\frac{n}{2}$ orthogonal hyperbolic planes, say P_i, $i=1, 2, \ .\ .\ ., \frac{n}{2}$. Associated with each P_i there is a basis of two isotropic vectors whose product is unity; let one of these vectors be designated h_i and the other designated h_{n+1-i}. When we run through the $\frac{n}{2}$ planes we obtain an ordered basis of n vectors, h_j, $j=1, 2, \ .\ .\ ., n$, with the special property that for each j the vector h_j has a unit product with h_{n+1-j} and is orthogonal to the remainder of the basis. If a matrix H is formed by letting h_j be the jth column of H then the following is obtained:

$$H^tH=\begin{pmatrix}
0 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 & 1\\
0 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 1 & 0\\
0 & 0 & 0 & 0 & \cdot & \cdot & 0 & 1 & 0 & 0\\
0 & 0 & 0 & 0 & \cdot & \cdot & 1 & 0 & 0 & 0\\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\
0 & 0 & 0 & 1 & \cdot & \cdot & 0 & 0 & 0 & 0\\
0 & 0 & 1 & 0 & \cdot & \cdot & 0 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 & 0\\
1 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 & 0
\end{pmatrix}$$

i.e., the Gram matrix H^tH has 1's on the second diagonal and 0's elsewhere.

If K is an $n \times n$ single Jordan block then the following identity is readily obtained:

$$(H^tH)\,K = K^t\,(H^tH).$$

Hence:

$$HKH^{-1} = (HKH^{-1})^t. \tag{6–6}$$

The matrix HKH^{-1} is symmetric and indecomposable.

Finally, assume that n is an odd integer and that there exists an $n-1$ dimensional hyperbolic subspace of F^n. Call this subspace U, and let U^* denote the space of vectors orthogonal to U. Since F^n is nonsingular it follows that:

$$\dim U + \dim U^* = \dim F^n = n.$$

Hence U^* is generated by a single vector k. The subspace U is also nonsingular since it is hyperbolic; consequently, U and U^* must be disjoint. Thus we can write:

$$F^n = U \perp k$$

Suppose the vector k has length square c; that is, $k^tk = c$. As in the first part of the proof we separate the space U into a system of orthogonal hyperbolic planes, P_i, $i = 1, 2, \ldots, \frac{(n-1)}{2}$. Associated with each plane there is a basis of two isotropic vectors whose product is unity; let one of these vectors be designated h_i, and multiply the second vector by the scalar c and designate the resultant vector by h_{n+1-i}. Finally define the vector $h_{(n+1)/2}$ to be the vector k, so that we obtain an ordered basis, h_j, $j = 1, 2, \ldots, n$, with the special property that for each j the vector h_j has a product with h_{n+1-j} equal to the scalar c, and is orthogonal to the remainder of the basis. The matrix H is defined to have the jth column equal to the basis vector h_j so that the Gram matrix H^tH consists of c's on the second diagonal and 0's elsewhere. Equation (6–6) continues to hold, and the matrix HKH^{-1} is symmetric and indecomposable.

To obtain a skew-symmetric indecomposable matrix we modify the matrix H used immediately above in the following way: negate columns $\frac{(n+1)}{2} - i$ where i runs through the odd integers less than $\frac{(n+1)}{2}$. Call this new matrix M. Then the Gram matrix M^tM consists of an alternation of $+c, -c, +c, -c, \ldots$ along the second diagonal and 0's elsewhere:

$$M^tM = \begin{pmatrix} 0 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 & \pm c \\ 0 & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & \mp c & 0 \\ 0 & 0 & 0 & 0 & \cdot & \cdot & 0 & \pm c & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdot & \cdot & \mp c & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \mp c & \cdot & \cdot & 0 & 0 & 0 & 0 \\ 0 & 0 & \pm c & 0 & \cdot & \cdot & 0 & 0 & 0 & 0 \\ 0 & \mp c & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 & 0 \\ \pm c & 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 & 0 \end{pmatrix}.$$

If K is a single Jordan block with 0's along the diagonal then the following holds:

$$(M^tM)K = -K^t(M^tM).$$

Hence the matrix MKM^{-1} is skew-symmetric indecomposable. *Q.E.D.*

COROLLARY: *The existence of any normal indecomposable matrix over* F *implies the existence of both symmetric and nonsymmetric normal indecomposable matrices of odd order over* F.

PROOF: Let B be the given normal indecomposable matrix. If B is of odd order then the existence of symmetric and nonsymmetric normal indecomposable matrices follows from the equivalence of conditions (2′), (3′) and (4′). If B is of even order, say $2n$, then F^{2n} must be hyperbolic. Consequently, F^{2n+1} contains a $2n$ dimensional hyperbolic subspace. The existence of symmetric and nonsymmetric normal indecomposable matrices of order $2n + 1$ now follows from the equivalence of conditions (1′), (2′), and (3′). *Q.E.D.*

If n is an even number we have shown that normal indecomposable matrices must be symmetric (Theorem 16), and that such matrices exist if and only if the underlying vector space is hyperbolic (Theorem 18).

If n is odd we have demonstrated that a normal matrix B with eigenvalue s must satisfy one of the following (Theorem 14):

(1) B is symmetric

(2) $(B-sI)^t = p(B-sI)$ for some polynomial p equivalent to the negative identity.

Furthermore there exist matrices satisfying (1) if and only if there exist matrices satisfying (2) if and only if the underlying vector space is the direct product of a one dimensional space and a hyperbolic space (Theorem 18).

It remains to show that condition (2) above is the most restrictive condition which can be placed on the nonsymmetric normal indecomposable matrices of odd order. This is proven in the following theorem.

THEOREM 19: *If for some polynomial equivalent to the negative identity, say* p(x), *there exists an indecomposable matrix with* 0 *eigenvalue,* A, *such that* $A^t = p(A)$, *then for any polynomial equivalent to the negative identity there exists such an associated matrix.*

PROOF: Let $q(x)$ be an arbitrary polynomial in the equivalence class of $-x$. Then $q(x)$ and $p(x)$ are equivalent, and there exists an admissible polynomial $r(x)$ such that:

$$p(x) = (r^{-1} \cdot q \cdot r)(x).$$

Evaluating at A we obtain:

$$A^t = p(A) = (r^{-1} \cdot q \cdot r)(A).$$

Or equivalently:

$$r(A^t) = (q \cdot r)(A).$$

The left side of the above equation is equal to $(r(A))^t$ and the right side is equal to $q(r(A))$. Hence we obtain:

$$B^t = q(B)$$

where $B = r(A)$. Since $r(x)$ is admissible the matrix B is indecomposable with 0 eigenvalue. *Q.E.D.*

The preceding discussion demonstrates that if there are any normal indecomposable matrices over F then there is an abundant supply of such matrices: the corollary to Theorem 18 proves the existence of nonsymmetric normal indecomposable matrices, and Theorem 19 proves the existence of the full spectrum, in terms of the transposing polynomials, of such matrices.

479-206 O - 73 - 4

On the basis of Theorem 18 we are also able to establish the existence of normal indecomposable matrices of higher order by considering the orthogonal sum of hyperbolic spaces. Suppose that there exist $n \times n$ symmetric indecomposable matrices and $m \times m$ symmetric indecomposable matrices over F where both n and m are even integers. Since F^n and F^m are hyperbolic, it follows that $F^n \perp F^m$ is hyperbolic. Since the space $F^n \perp F^m$ is isomorphic to F^{n+m} it follows that there must exist $(n+m) \times (n+m)$ symmetric indecomposable matrices over F. If the integer n is even but m is odd then the vector space $F^n \perp F^m$ must contain a hyperbolic subspace of dimension $n+m-1$, and consequently there must exist normal indecomposable matrices of this order.

Of particular interest are those fields which contain 2-dimensional isotropic vectors. Examples of such fields are $GF(5)$ and the complex numbers (with, of course, identity automorphism). Since there exists a 2-dimensional isotropic vector there must exist 2×2 symmetric indecomposable matrices, for if (a, b) is isotropic then such a matrix is given by the following:

$$S = \begin{pmatrix} a & b \\ b & -a \end{pmatrix} = \begin{pmatrix} a & 1 \\ b & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & 1 \\ b & 0 \end{pmatrix}^{-1}.$$

Consequently, there must also exist 3×3 normal indecomposable matrices. Thus the vector space F^2 is hyperbolic and F^3 contains a 2-dimensional hyperbolic subspace. Now let n represent an arbitrary integer greater than 1. If n is even then the vector space F^n is hyperbolic since it is expressible in the form $F^2 \perp F^2 \perp \ldots \perp F^2$; consequently, there exist $n \times n$ normal indecomposable matrices over F. If, on the other hand, n is odd then the vector space F^n contains an $n-1$ dimensional hyperbolic subspace since it is expressible in the form $F^2 \perp F^2 \perp \ldots \perp F^2 \perp F$; consequently, there again exist $n \times n$ normal indecomposable matrices over the given field. Since any matrix of order 1 is normal indecomposable the following holds: If the base field has a 2-dimensional isotropic vector then there exist normal indecomposable matrices of all orders.

It should be noted that while the existence of normal indecomposable matrices of order $2n$ implies the existence of such matrices of order $2n+1$, the converse is not true. As a counterexample let F be a field for which there exists 3-dimensional isotropic vectors but there do not exist 2-dimensional isotropic vectors (e.g., $GF(7)$). Since there are no 2-dimensional isotropic vectors over F, the space of 2-dimensional vectors is not hyperbolic, and according to Theorem 18 there exist no 2×2 normal indecomposable matrices. However, since there do exist 3-dimensional isotropic vectors, the space of 3-dimensional vectors does contain a hyperbolic plane. Consequently, there must exist 3×3 normal indecomposable matrices (Theorem 18). Example 16 is a 3×3 normal indecomposable matrix over $GF(7)$.

EXAMPLE 16: Normal indecomposable matrix of order 3 where no such matrix of order 2 exists. Let the base field be $GF(7)$.

$$\begin{pmatrix} 0 & 3 & 5 \\ 4 & 0 & 6 \\ 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 0 \\ 1 & 1 & 0 \\ 5 & 0 & 6 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 2 & 0 \\ 1 & 1 & 0 \\ 5 & 0 & 6 \end{pmatrix}^{-1}.$$

While the existence of normal indecomposable matrices of even order imply the existence of such matrices of the next highest odd order, we are unable to go from odd order to the next highest even order. In other words, given that there exist normal indecomposable matrices of order n, it is not necessarily true that there exist such matrices of order m for $m > n$. As a counterexample we can again consider $GF(7)$. The vector space F^4 contains the following two independent, isotropic, orthogonal vectors

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 4 \\ 2 \\ 1 \end{pmatrix}.$$

Consequently F^4 is hyperbolic and there exist 4×4 and 5×5 indecomposable normal matrices. However, there do not exist any 6×6 normal indecomposable matrices over $GF(7)$. The proof of this is lengthy and has been put in section 7. Since F^7 can be expressed as the orthogonal sum of F^4 (which is hyperbolic) and F^3 (which contains a 2-dimensional hyperbolic subspace), it follows that F^7 contains a 6-dimensional hyperbolic subspace; consequently, there do exist 7×7 normal indecomposable matrices over this field.

7. Proof of Nonexistence of 6×6 Normal Indecomposable Matrices over $GF(7)$

If there exist normal indecomposable 6×6 matrices over $GF(7)$ then $GF(7)^6$ is hyperbolic and contains three independent, isotropic, orthogonal vectors. The following modifications of such a triplet of vectors does not affect the independence, orthogonality, or isotropicity:

(1) Addition of a multiple of one vector to a second vector.
(2) Permutation of the elements of each vector, always using the same permutation.
(3) Multiplication of any vector by a nonzero constant.

Using modification of types (1) and (2), the existence of any such triplet ensures the existence of three isotropic, orthogonal vectors having the following form:

$$\begin{pmatrix} x_1 \\ 0 \\ 0 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} \begin{pmatrix} 0 \\ y_2 \\ 0 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \end{pmatrix} \tag{7-1}$$

where x_1, y_2, and z_3 are nonzero. Since each of the three vectors are nonzero and since the spaces of one and two dimensional vectors over $GF(7)$ contain no isotropic vectors, each of these vectors must contain at least three nonzero elements.

LEMMA 7: *At most one vector of the triplet (7–1) contains exactly three nonzero elements.*

PROOF: Assume $x_6=0$, and suppose y_6 also equals zero. Since $x\cdot z=y\cdot z=0$ the following homogeneous equations hold:

$$x_4\cdot z_4+x_5\cdot z_5=0$$

$$y_4\cdot z_4+y_5\cdot z_5=0.$$

Consequently, (x_4, x_5) and (y_4, y_5) are linearly dependent (z contains at least three nonzero elements so that z_4 and z_5 cannot be simultaneously zero). Thus:

$$0=x\cdot y=s(x_4^2+x_5^2)$$

for some scalar s. But this is impossible since there are no 2-dimensional isotropic vectors over this base field. Hence $y_6\neq0$. Similarly $z_6\neq0$.

Since $x\cdot y=x_4\cdot y_4+x_5\cdot y_5=0$ where $x_4\neq0$, $x_5\neq0$, and since y_4 and y_5 cannot be simultaneously zero, it follows that y_4 and y_5 must be nonzero. Similarly, z_4 and z_5 are nonzero. *Q.E.D.*

LEMMA 8: *No vector of the triplet (7–1) has exactly three nonzero elements.*

PROOF: Assume $x_6=0$. Since $x\cdot y=x\cdot z=0$ the following homogeneous equations hold:

$$x_4\cdot y_4+x_5\cdot y_5=0$$

$$x_4\cdot z_4+x_5\cdot z_5=0.$$

Hence (y_4, y_5) and (z_4, z_5) must be dependent. We can thus consider the three vectors (7–1) to be in the following normalized form:

$$\begin{pmatrix} x_1 \\ 0 \\ 0 \\ 1 \\ x_5 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ y_2 \\ 0 \\ 1 \\ -\frac{1}{x_5} \\ y_6 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ z_3 \\ 1 \\ -\frac{1}{x_5} \\ z_6 \end{pmatrix} \tag{7–2}$$

where the orthogonality of x and y and of x and z have been explicitly accounted for.

Since the vector $\mathbf{x}$ is isotropic the following holds:

$$x_1^2+1+x_5^2=0.$$

Hence:

$$x_5^2=6-x_1^2.$$

Since the only squares over $GF(7)$ are 1, 2, 4 it follows that x_5^2 must assume the value 2 or 4.

From the orthogonality of y and z we obtain:

$$1+\frac{1}{x_5^2}+y_6 \cdot z_6=0.$$

Substituting the two possible values of x_5^2 we obtain:

$$y_6 \cdot z_6 = x_5^2.$$

CASE 1: $x_5^2=2$.

Since y and z are isotropic the following equations hold:

$$y_2^2+1+4+y_6^2=0$$

$$z_3^2+1+4+z_6^2=0.$$

Hence:

$$y_2^2=2-y_6^2 \neq 0.$$

$$z_3^2=2-z_6^2 \neq 0.$$

Again using the fact that the only squares are 1, 2, 4 it follows that $y_6^2=z_6^2=1$. Hence $y_6 \cdot z_6=\pm 1=x_5^2$. Contradiction.

CASE 2: $x_5^2=4$.

In a similar way we obtain:

$$y_2^2=4-y_6^2 \neq 0$$

$$z_3^2=4-z_6^2 \neq 0.$$

Hence $y_6^2=z_6^2=2$, and as a result: $y_6 \cdot z_6=\pm 2 \neq x_5^2$. Contradiction. *Q.E.D.*

As a result of the above lemma if there are to exist three independent, isotropic, orthogonal vectors over $GF(7)$, they can be assumed to be of the form (7–1) where all terms written with letters are, in fact, nonzero. Assume that these vectors are normalized so that $x_1=y_2=z_3=2$. Then the following vectors are orthogonal, and have length 3:

$$\begin{pmatrix} x_4 \\ x_5 \\ x_6 \end{pmatrix}, \begin{pmatrix} y_4 \\ y_5 \\ y_6 \end{pmatrix}, \begin{pmatrix} z_4 \\ z_5 \\ z_6 \end{pmatrix}.$$

The following is a complete list of 3-dimensional vectors over $GF(7)$ having length 3:

$$\begin{pmatrix} \pm 1 \\ \pm 1 \\ \pm 1 \end{pmatrix}, \begin{pmatrix} \pm 2 \\ \pm 2 \\ \pm 3 \end{pmatrix}, \begin{pmatrix} \pm 2 \\ \pm 3 \\ \pm 2 \end{pmatrix}, \begin{pmatrix} \pm 3 \\ \pm 2 \\ \pm 2 \end{pmatrix}.$$

It can be shown that there do not exist three orthogonal vectors in this set.

Hence there do not exist three orthogonal, isotropic, independent vectors over $GF(7)^6$, and consequently there do not exist any 6×6 normal indecomposable matrices over $GF(7)$.

8. Some Open Questions

1. Identify the group of the jth order admissible polynomials with entries from the field $GF(n^k)$.

2. Find the polynomials conjugate to $-x$. It can be shown that if

$$p(x)=p_{n-1}x^{n-1}+\ldots+p_1x$$

is conjugate to $-x$ then

$$\begin{aligned} p_1&=-1, \\ p_3&=-p_2^2, \\ p_5&=2p_2^4-3p_2p_4, \\ p_7&=-13p_2^6+18p_2^3p_4-4p_2p_6. \end{aligned}$$

We conjecture that all odd numbered coefficients can be expressed in terms of the preceding even numbered coefficients.

3. Characterize the normal matrices over an arbitrary field of characteristic $\neq 2$ which are expressable as a polynomial in their conjugate transpose.

4. Characterize the normal matrices over an arbitrary field of characteristic $\neq 2$ which are neither unitarily similar to a diagonal matrix nor have a single elementary divisor.

9. References

[1] Artin, E., Geometric Algebra (Interscience, New York, 1957).
[2] Drazin, M. P., On diagonable and normal matrices, Quart. J. Math. (Oxford) **2**, 189–198 (1951).
[3] Gantmacher, F. R., The Theory of Matrices (Chelsea, New York, 1959).
[4] Hoffman, A. J., and Taussky, O., A characterization of Normal Matrices, J. Res. Nat. Bur. Stand. (U.S.), **52**, No. 1, 17–19 (1954).
[5] Jacobson, N., Lectures in Abstract Algebra, Vol. **II**, (Van Nostrand, Princeton, 1953).
[6] Katz, I. J., and Pearl, M., On *EPr* and Normal *EPr* Matrices, J. Res. Nat. Bur. Stand. (U.S.), **70B** (Math. Phys.), No. 1, 47–77 (1966). RP 2467.
[7] Katz, I. J., and Pearl, M., Solutions of the Matrix Equation $A^*=XA=AX$, J. London Math. Soc. **41**, 443–452 (1966).
[8] Lang, S., Algebra (Addison Wesley, Reading, 1965).
[9] Pearl, M., On normal and *EPr* matrices, Mich. Math. J. **6**, 1–5 (1959).
[10] Pearl, M., On normal *EPr* matrices, Mich. Math. J. **8**, 33–37 (1961).
[11] Perlis, S., Theory of Matrices (Addison Wesley, Reading, 1952).
[12] Schwerdtfeger, H., Introduction to Linear Algebra and the Theory of Matrices (P. Noorhoff, Groningen, 1961).
[13] Toeplitz, O., Das algebraische Analogen zu einen Satz von Fejer, Math. Zeitschrift. **2**, 187–197 (1918).
[14] Williamson, J., Matrices normal with respect to an Hermitian matrix, Amer. J. Math. **60**, 355–373 (1938).
[15] Williamson, J., Normal matrices over an arbitrary field of characteristic zero, Amer. J. Math. **61**, 335–356 (1939).

(Paper 76B3&4–366)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Vol. 76B, Nos. 3 and 4, July–December 1972

Distance Coordinates With Respect to a Triangle of Reference

Karl Goldberg

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

(July 21, 1972)

With respect to a triangle of reference $A_1A_2A_3$, each point P in the plane of the triangle, has unique *area* coordinates: $P = (b_1, b_2, b_3)$ with $b_1 + b_2 + b_3 = 1$. *Distance* coordinates are introduced such that $P = [d_1, d_2, d_3]$, with d_k the distance from P to A_k. It is shown that there is an explicit function $f(x_1, x_2, x_3)$ such that $f(d_1^2, d_2^2, d_3^2) = 0$ is necessary and sufficient for $P = [d_1, d_2, d_3]$, each d_k nonnegative. The partial derivatives $f_k(x_1, x_2, x_3) = \partial f(x_1, x_2, x_3)/\partial x_k$ are such that $b_k = f_k(d_1^2, d_2^2, d_3^2)$ for each k. Other results relating the b_k and the d_k are given. The use of $f(x_1, x_2, x_3)$ in solving geometric problems is shown.

Key words: Area coordinates; distance coordinates; Plane Geometry; radical center; triangle of reference.

We are given three noncollinear points A_1, A_2, A_3 and all other points are in the plane of the triangle of reference $A_1A_2A_3$.

Notationally, two distinct points X, Y determine an infinite line XY, with the finite line segment $\overline{XY}$ having length $|X, Y|$. If X, Y are centers of circles with radii x, y respectively, the *radical axis* of those circles is a line perpendicular to XY, at a point which is $(|X, Y|^2 + x^2 - y^2)/2|X, Y|$ from X in the direction of Y. Given a third point Z, not on XY, as the center of a circle, the three radical axes meet in a common point called their *radical center*. The area of the triangle XYZ is denoted by $|X, Y, Z|$. The function

$$F(x_1, x_2, x_3) = 2(x_1x_2 + x_1x_3 + x_2x_3) - x_1^2 - x_2^2 - x_3^2 \tag{1}$$

has the well-known property

$$16|X, Y, Z|^2 = F(|X, Y|^2, |X, Z|^2, |Y, Z|^2). \tag{2}$$

Let $\Delta = |A_1, A_2, A_3|$ denote the area of the triangle of reference. With respect to this triangle every point P has unique *area coordinates*[1] b_1, b_2, b_3, which are real numbers restricted by

$$b_1 + b_2 + b_3 = 1. \tag{3}$$

We write

$$P = (b_1, b_2, b_3),$$

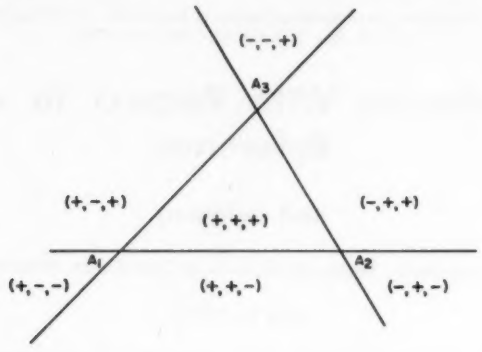
with $A_1 = (1, 0, 0), A_2 = (0, 1, 0), A_3 = (0, 0, 1)$.

AMS Subject Classification: Primary 5010.

[1] Also called "normalized barycentric" or "areal" coordinates [1].[2]

[2] Coxeter, H.S.M., Introduction to Geometry (Wiley, 1961).

The area coordinate b_1 is defined as the ratio $\pm|P, A_2, A_3|/\Delta$ with $b_1 > 0$ if $\overline{A_1P}$ does not intersect A_2A_3, $b_1 \leq 0$ if it does. Similarly for the other b_k, so that the diagram of signs is



Conversely, any three real numbers b_1, b_2, b_3, restricted by (3), define a unique point $P=(b_1, b_2, b_3)$ with respect to $A_1A_2A_3$.

If a point P is at a distance d_1 from A_1, d_2 from A_2, and d_3 from A_3, we call these the *distance coordinates* of P with respect to $A_1A_2A_3$, and write

$$P=[d_1, d_2, d_3],$$

or, when more convenient,

$$P=<d_1^2, d_2^2, d_3^2>.$$

For clarity, $d_k=|P, A_k|$ *for all* $k=1, 2, 3$.

Note that the triangle PA_2A_3 has side lengths a_1, d_2, d_3 so that

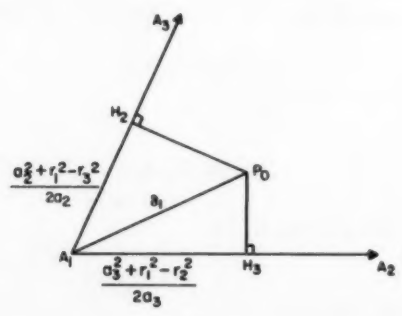
$$16\Delta^2 b_1^2 = F(a_1^2, d_2^2, d_3^2) \tag{4}$$

with similar equations for the other b_k. Also note that P is the radical center of circles with centers A_1, A_2, A_3 and radii d_1, d_2, d_3 respectively.

Now consider the general case of circles with centers A_1, A_2, A_3 and radii r_1, r_2, r_3. Denote their radical center by

$$P_0=[\delta_1, \delta_2, \delta_3]. \tag{5}$$

In what follows there is no loss of generality in assuming that P_0 is in the interior of $A_1A_2A_3$. Near A_1 we have



Let α_1 denote the interior angle at A_1. Then $|H_2, H_3| = \delta_1 \sin \alpha_1$.[3] The formula for $\cos \alpha_1$ in the triangle $A_1H_2H_3$ is

$$(6) \qquad 2\left(\frac{a_2^2+r_1^2-r_3^2}{2a_2}\right)\left(\frac{a_3^2+r_1^2-r_2^2}{2a_3}\right)\cos\alpha_1 = \left(\frac{a_2^2+r_1^2-r_3^2}{2a_2}\right)^2 + \left(\frac{a_3^2+r_1^2-r_2^2}{2a_3}\right)^2 - (\delta_1 \sin\alpha_1)^2.$$

In order to simplify this equation we shall need a few definitions and formulas. Define

$$(7) \qquad \begin{aligned} c_1 &= a_2^2 + a_3^2 - a_1^2 \\ c_2 &= a_1^2 + a_3^2 - a_2^2 \\ c_3 &= a_1^2 + a_2^2 - a_3^2. \end{aligned}$$

Note that

$$(8) \qquad c_1 = 2a_2^2 - c_3 = 2a_3^2 - c_2$$

and that

$$(9) \qquad 2a_2a_3 \cos\alpha_1 = c_1.$$

Since

$$(10) \qquad a_2a_3 \sin\alpha_1 = 2\Delta,$$

we have

$$(11) \qquad c_1^2 + 16\Delta^2 = 4a_2^2a_3^2.$$

Now return to eq (6). Multiply through by $4a_2^2a_3^2$, and use eqs (7) through (11) to simplify. We get $16\Delta^2 f(r_1^2, r_2^2, r_3^2) = 16\Delta^2(r_1^2 - \delta_1^2)$, where

$$(12) \qquad 16\Delta^2 f(x_1, x_2, x_3) = \sum_{k=1}^{3} a_k^2 c_k x_k - a_1^2a_2^2a_3^2 - \frac{1}{2}\{c_1(x_2-x_3)^2 + c_2(x_1-x_3)^2 + c_3(x_1-x_2)^2\}.$$

Generally:

$$(13) \qquad f(r_1^2, r_2^2, r_3^2) = r_k^2 - \delta_k^2. \qquad k = 1, 2, 3.$$

If $f(r_1^2, r_2^2, r_3^2) = 0$ then $r_k = \delta_k$ for all $k = 1, 2, 3$ so that $P_0 = [r_1, r_2, r_3]$.
Now notice that

$$(14) \qquad \sum_{k=1}^{3} a_k^2 c_k = F(a_1^2, a_2^2, a_3^2) = 16\Delta^2,$$

[3] The circle with diameter $\overline{A_1P_0}$ goes through H_2 and H_3. Thus $\overline{H_2H_3}$ is a chord with opposite angle α_1.

so that

$$f(x_1 - t, x_2 - t, x_3 - t) = f(x_1, x_2, x_3) - t \qquad \text{all } t. \tag{15}$$

Let $x_k = r_k^2$ for all $k = 1, 2, 3$ and let $t = f(r_1^2, r_2^2, r_3^2)$. Since (13) implies $x_k - t = \delta_k^2$ for all $k = 1, 2, 3$, we have $f(\delta_1^2, \delta_2^2, \delta_3^2) = 0$. As we pointed out before, any point $P = [d_1, d_2, d_3]$ can be a radical center. The above shows that $f(d_1^2, d_2^2, d_3^2) = 0$.

Combining this with our remarks following eq (13), we have proved the first part of the following theorem:

THEOREM. *Let* d_1, d_2, d_3 *be any real nonnegative numbers. Then there is a point* $P = [d_1, d_2, d_3]$ *if and only if* $f(d_1^2, d_2^2, d_3^2) = 0$ (f *defined in* (12)). *In that case* $P = (f_1(d_1^2, d_2^2, d_3^2), f_2(d_1^2, d_2^2, d_3^2), f_3(d_1^2, d_2^2, d_3^2))$ *where* $f_k(x_1, x_2, x_3) = \partial f(x_1, x_2, x_3)/\partial x_k$ *for all* $k = 1, 2, 3$.

For clarity, we write out the f_k:

$$\begin{aligned} 16\Delta^2 f_1(x_1, x_2, x_3) &= a_1^2 c_1 + c_3(x_2 - x_1) + c_2(x_3 - x_1) \\ 16\Delta^2 f_2(x_1, x_2, x_3) &= a_2^2 c_2 + c_3(x_1 - x_2) + c_1(x_3 - x_2) \\ 16\Delta^2 f_3(x_1, x_2, x_3) &= a_3^2 c_3 + c_2(x_1 - x_3) + c_1(x_2 - x_3). \end{aligned} \tag{16}$$

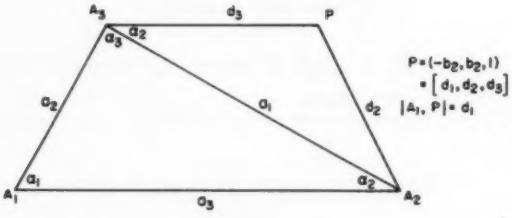
Note that (14) implies

$$f_1(x_1, x_2, x_3) + f_2(x_1, x_2, x_3) + f_3(x_1, x_2, x_3) = 1. \tag{17}$$

Also note that

$$f_k(x_1 - t, x_2 - t, x_3 - t) = f_k(x_1, x_2, x_3) \qquad \text{all } t;\ k = 1, 2, 3. \tag{18}$$

In the process of proving the last part of the theorem, we shall need the following computations. The locus of points with $b_3 = 1$ is a line through A_3 parallel to A_1A_2. Naturally if $P = (b_1, b_2, b_3)$ then $b_1 + b_2 = 0$ (see (3)). Suppose $b_1 \leqslant 0$.



The area of PA_2A_3 is $\Delta \cdot |-b_2|$. It is also $\frac{1}{2} a_1 d_3 \sin \alpha_2$.

Therefore

$$d_3 = a_3 b_2. \tag{19}$$

Using the cosine formula in that triangle yields

$$\begin{aligned} d_2^2 &= d_3^2 + a_1^2 - 2a_1 d_3 \cos \alpha_2 \\ &= a_3^2 b_2^2 + a_2^2 - b_2 c_2. \end{aligned} \tag{20}$$

Using the cosine formula in the triangle A_1A_3P we have

$$d_1^2 = d_3^2 + a_2^2 - 2a_2 d_3 \cos(\alpha_2 + \alpha_3)$$

(21) $$= a_3^2 b_2^2 + a_2^2 + b_2 c_1.$$

Now use eq (18) with each $x_k = d_k^2$ and $t = d_3^2 = a_3^2 b_2^2$. For $k = 1$ we have (using (20) and (21)):

$$16\Delta^2 f_1(d_1^2, d_2^2, d_3^2) = 16\Delta^2 f_1(a_2^2 + b_2 c_1, a_1^2 - b_2 c_2, 0)$$

$$= a_1^2 c_1 + c_3(a_1^2 - a_2^2 - b_2(c_1 + c_2)) - c_2(a_2^2 + b_2 c_1)$$

(using (16)). Since $a_1^2 c_1 + c_3(a_1^2 - a_2^2) = c_2 a_2^2$ (note $c_1 + c_3 = 2a_2^2$, $c_2 + c_3 = 2a_1^2$), and since

(22) $$c_1 c_2 + c_1 c_3 + c_2 c_3 = 16\Delta^2,$$

($c_1 + c_2 = 2a_3^2$, then add the three symmetric formulas, and use eq (14)), we have

$$f_1(d_1^2, d_2^2, d_3^2) = -b_2 = b_1$$

as desired.

Similar computations prove $f_2(d_1^2, d_2^2, d_3^2) = b_2$ and (as a check), $f_3\ (d_1^2, d_2^2, d_3^2) = 1 = b_3$. The case $b_2 \leq 0$ is handled similarly to prove the last part of the theorem for the case $b_3 = 1$.

Return to eq (12) and solve $f(x_1, x_2, x_3) = 0$ for x_1. An intermediary stage is the equation

(23) $$16\Delta^2\{f_1(x_1, x_2, x_3)\}^2 = 2a_1^2(x_2 + x_3) + 2x_2 x_3 - a_1^4 - x_2^2 - x_3^2.$$

The r.h.s. is recognized to be $F(a_1^2, x_2, x_3)$. Let $x_k = d_k^2$ for all $k = 1,2,3$, and use (4), to get $\{f_1(d_1^2, d_2^2, d_3^2)\}^2 = b_1^2$. Generally

$$f_k(d_1^2, d_2^2, d_3^2) = \pm b_k \qquad k = 1,2,3$$

Set $f_k = f_k(d_1^2, d_2^2, d_3^2)$ for all $k = 1, 2, 3$. Equations (3) and (17), showing $\Sigma b_k = \Sigma f_k = 1$, imply that we cannot have $f_k = -b_k$ for all k. Suppose $f_3 = b_3$. Then

$$f_1 + f_2 = 1 - f_3 = 1 - b_3 = b_1 + b_2.$$

If $f_1 = b_1$ then $f_2 = b_2$, and conversely. The only open case is $f_1 = -b_1$, $f_2 = -b_2$. This implies $b_1 + b_2 = 0$, whence $b_3 = 1$. We have already covered this case, so the proof of the theorem is complete.

An interesting implication for P_0 is immediate. Use eq (18) with $x_k = r_k^2$ for all $k = 1, 2, 3$ and $t = f(r_1^2, r_2^2, r_3^2)$ as before. The result is $f_k(\delta_1^2, \delta_2^2, \delta_3^2) = f_k(r_1^2, r_2^2, r_3^2)$ for all $k = 1, 2, 3$. In other words the area coordinates for the radical center of three circles with centers at A_1, A_2, A_3 and radii r_1, r_2, r_3 respectively are given by

(24) $$P_0 = (f_1(r_1^2, r_2^2, r_3^2), f_2(r_1^2, r_2^2, r_3^2), f_3(r_1^2, r_2^2, r_3^2)).$$

Of course, the distance coordinates are given by

(25) $$P_0 = \langle r_1^2 - f(r_1^2, r_2^2, r_3^2), r_2^2 - f(r_1^2, r_2^2, r_3^2), r_3^2 - f(r_1^2, r_2^2, r_3^2) \rangle.$$

If we wish to find the (0 to 8) circles simultaneously tangent to the three circles used above, we can do so through f, to obtain four quadratic equations whose solutions solve the problem. The point is that a circle of radius r which is simultaneously tangent to all three circles has a center

(26) $$P=[r_1+\epsilon_1 r,\ r_2+\epsilon_2 r,\ r_3+\epsilon_3 r],$$

with each $\epsilon_k=\pm 1$ depending on whether the tangency is "outside" ($\epsilon=1$) or "inside" ($\epsilon=-1$). Simplifying

(27) $$f((r_1+\epsilon_1 r)^2,\ (r_2+\epsilon_2 r)^2,\ (r_3+\epsilon_3 r)^2)=0,$$

we get a quadratic equation in r (with constant term $f(r_1^2,\ r_2^2,\ r_3^2)$). If r is a negative root of that equation, we simply "assign" $-r$ to $-\epsilon_1, -\epsilon_2, -\epsilon_3$ since $\epsilon_k r=(-\epsilon_k)\ (-r)$, $k=1, 2, 3$.

Thus we can cover all solutions with just four triples of ϵ's, no two of which are negatives of each other.

We end this note with a list of formulas connecting the area and distance coordinates of a point

$$P=(b_1,\ b_2,\ b_3)=[d_1,\ d_2,\ d_3].$$

The formulas are given without proof, but are easily derived, with extensive use of the formula for the distance between P and $P'=(b_1',\ b_2',\ b_3')$:

(28) $$2\,|P,P'\,|^2=\sum_{k=1}^{3} c_k(b_k-b_k')^2.$$

First we complete the connection between the coordinates, begun in the last formula of the theorem, with

$$d_1^2=a_3^2b_2^2+c_1b_2b_3+a_2^2b_3^2$$

(29) $$d_2^2=a_3^2b_1^2+c_2b_1b_3+a_1^2b_3^2$$

$$d_3^2=a_2^2b_1^2+c_3b_1b_2+a_1^2b_2^2$$

or

(30) $$2d_k^2=(1-2b_k)c_k+\sum_{n=1}^{3} b_n^2c_n \qquad k=1,2,3.$$

Let R denote the circumradius of $A_1A_2A_3$, and ρ_P the distance from P to the circumcenter. The latter has area coordinates $a_k^2c_k/16\Delta^2$, $k=1,2,3$. Also $4\Delta R=a_1a_2a_3$.

Define

(31) $$G_P=R^2-\rho_P^2.$$

Then G_P can be found using only the b_k:

(32) $$G_P=a_1^2b_2b_3+a_2^2b_1b_3+a_3^2b_1b_2$$

(33) $$=\frac{1}{2}\sum_{k=1}^{3}(b_k-b_k^2)c_k;$$

or only the d_k:

(34) $$G_P=R^2-(c_1(d_2-d_3)^2+c_2(d_1-d_3)^2+c_3(d_1-d_2)^2)/32\Delta^2$$

$$= R^2 - (a_1^2(d_1^2 - d_2^2)(d_1^2 - d_3^2) + a_2^2(d_2^2 - d_1^2)(d_2^2 - d_3^2) + a_3^2(d_3^2 - d_1^2)(d_3^2 - d_2^2))/16\Delta^2 \tag{35}$$

$$= (2a_1^2a_2^2a_3^2 - \sum a_k^2c_kd_k^2)/16\Delta^2; \tag{36}$$

or symmetric combinations:

$$G_P = \sum_{k=1}^{3} b_kd_k^2 \tag{37}$$

$$= \sum_{k=1}^{3} (a_k^2 - 2d_k^2 + b_k^2c_k)/4 \tag{38}$$

$$= \sum_{k=1}^{3} (a_k^2 - 2d_k^2 + b_kc_k)/6 \tag{39}$$

$$= \sum_{k=1}^{3} (a_k^2 - d_k^2 - b_ka_k^2)/3 \tag{40}$$

$$= 2\left\{b_1a_2^2a_3^2 + b_2a_1^2a_3^2 + b_3a_1^2a_2^2 - \sum_{k=1}^{3} a_k^2d_k^2\right\}\Big/\sum_{k=1}^{3} a_k^2 \tag{41}$$

$$= \left\{b_1b_2b_3\sum_{k=1}^{3} a_k^2 + \sum_{k=1}^{3} b_k^2d_k^2\right\}\Big/2(b_1b_2 + b_1b_3 + b_2b_3) \tag{42}$$

(except at a vertex). Other relations are

$$d_1^2 = b_2a_3^2 + b_3a_2^2 - G_P \qquad \text{etc.} \tag{43}$$

$$b_1c_1 = 2G_P - a_1^2 + d_2^2 + d_3^2 \qquad \text{etc.} \tag{44}$$

$$(1 - b_1)G_P = b_1d_1^2 + b_2b_3a_1^2 \qquad \text{etc.} \tag{45}$$

$$2a_2^2a_3^2b_1 = c_1G_P + a_2^2d_2^2 + a_3^2d_3^2 - a_1^2d_1^2 \qquad \text{etc.} \tag{46}$$

The pedal triangle of P, which has side lengths $a_kd_k/2R$, $k = 1, 2, 3$, *has area* $\Delta|G_P|/4R^2$; i.e.,

$$16\Delta^2G_P^2 = F(a_1^2d_1^2, a_2^2d_2^2, a_3^2d_3^2) \tag{47}$$

$$= 4a_2^2a_3^2d_2^2d_3^2 - (a_2^2d_2^2 + a_3^2d_3^2 - a_1^2d_1^2)^2 \qquad \text{etc.} \tag{48}$$

If P does not lie on the triangle $A_1A_2A_3$ then the lines A_1P, A_2P, A_3P intersect the circle of radius ρ_P, concentric with the circumcircle, in P and in points A_1', A_2', A_3' respectively which have (opposite) side lengths $\lambda_P|b_k|d_k$ for $k = 1, 2, 3$; $\lambda_P = 4\Delta\rho_P/d_1d_2d_3$. The area of $A_1'A_2'A_3'$ is $\Delta\lambda_P^2|b_1b_2b_3|$. Thus

$$16\Delta^2(b_1b_2b_3)^2 = F(b_1^2d_1^2, b_2^2d_2^2, b_3^2d_3^2). \tag{49}$$

Finally, suppose $d_k^2 = g_k(t)$, a differentiable function of t, for $k = 1, 2, 3$. The b_k will also be differentiable functions of t, (by the last statement of the theorem), and b_k' will denote the derivative. We have

(50) $$\sum_{k=1}^{3} b_k g_k'(t) = 0$$

and

(51) $$g_1'(t) + c_1 b_1' = g_2'(t) + c_2 b_2' = g_3'(t) + c_3 b_3' = \sum_{k=1}^{3} c_k b_k b_k'.$$

(Paper 76B3&4–367)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Vol. 76B, Nos. 3 and 4, July–December 1972

On Regular Sets of Polynomials Whose Zeros Lie in Prescribed Domains

M. A. Farag* and M. Nassif*

(July 18, 1972)

The relation between the mode of increase of regular basic sets of finite span and the order of magnitude of the zeros of polynomials $\{p_n(z)\}$ belonging to them is investigated. Upper bounds are obtained for the order of the basic sets when the zeros of $p_n(z)$ lie either in the unit circle or in a circle whose radius increases in a certain manner with the index n of the polynomial.

Key words: Basic sets; Cannon sum of basic sets; lower semiblock matrices; order of basic sets; zeros of polynomials in regular sets.

1. Introduction

The relation between the mode of increase of simple sets[1] and the order of magnitude of the zeros of polynomials belonging to them has been the interest of many authors, of whom we may mention Eweida [1][2] and Nassif [3]. The same problem is studied in the present paper when the sets considered are *regular of finite span*. To formulate a precise definition of such sets we suppose that l is an integer greater than 1,[3] and the sequence (μ_n) of integers is constructed so that

$$\mu_0=0,\ 1\leq \mu_n-\mu_{n-1}\leq l \quad ;\ (n\geq 1). \tag{1.1}$$

Thus, if we put

$$\nu_n=\mu_n-\mu_{n-1} \quad ;\ (n\geq 1), \tag{1.2}$$

then

$$1\leq \nu_n\leq l \quad ;\ n\leq \mu_n\leq nl. \tag{1.3}$$

Let $\{p_n(z)\}$ be a set of polynomials and let d_n be the degree of the polynomial $p_n(z)$, so that we can write

$$p_n(z)=\sum_{k=0}^{d_n} p_{n,k}z^k. \tag{1.4}$$

We shall assume that

$$d_0=0;\ m<d_m\leq \mu_n \quad \text{when} \quad \mu_{n-1}<m<\mu_n;\ d\mu_n=\mu_n;\ (n\geq 1). \tag{1.5}$$

AMS Subject Classification Numbers: Primary 3010.

*An invited paper. Present address, Mathematics Department, Assiut University, Assiut, Egypt.

[1] The reader is supposed to be acquainted with the theory of basic sets of polynomials as given by Whittaker [4, 5]. The mode of increase of a basic set is determined by its order and type, c. f. Whittaker [4; pp. 11, 12]. Our concern here is with the order.

[2] Figures in brackets indicate the literature references at the end of this paper.

[3] When $l=1$, the set will be simple.

When the set $\{p_n(z)\}$ thus defined, is basic it is said to be *regular of finite span of bound l*, and the set is normalized in the sense that

$$p_{n,d_n}=1 \qquad ;\ (n\geq 0). \tag{1.6}$$

The matrix $(\rho_{n,k})$ of coefficients of such sets is a lower semi-block[4] matrix whose leading diagonal consists of square matrices (Δ_n) of the form

$$\begin{cases}\Delta_0=(1)\\ \Delta_n=(p_{\mu_n-r,\mu_n-s});\ 0\leq r,\ s\leq \nu_n-1;\ (n\geq 1).\end{cases} \tag{1.7}$$

Let δ_n denote the modulus of the determinant $|\Delta_n|$; δ_n should be positive to ensure that the set is basic. Write

$$\sigma=\lim_{n\to\infty}\inf\frac{\log \delta_1\delta_2\ldots\delta_n}{n\log n}. \tag{1.8}$$

With the above notation, the main results of the present work are formulated in the following theorems.[5]

THEOREM 1: *Let* $\{p_n(z)\}$ *be a regular set of polynomials of finite span of bound l, and suppose that the zeros of the polynomials* $\{p_n(z)\}$ *all lie in* $|z|\leq 1$. *Then the set will be of order not exceeding* $\frac{1}{2}(l+1)-\frac{\sigma}{l}$, *and this bound is attainable.*

THEOREM 2: *Let the set* $\{p_n(z)\}$ *be as in Theorem* 1 *and suppose that the zeros of the polynomial* $p_n(z)$ *lie in* $|z|\leq n^{\alpha}$, *where* α *is a positive finite number. Then the order of the set* $\{p_n(z)\}$ *will not exceed* $\frac{1}{2}(\alpha+1)(l+1)-\frac{\sigma}{l}$.

2. Preliminary Results

We shall establish in this section a lemma, of general type, which is the basis for the proofs of the above theorems. In fact, we shall suppose that the zeros of the polynomial $p_n(z)$ lie in $|z|\leq\rho_n$, where the numbers (ρ_n) accord to the following restrictions.

$$\rho_{n+1}\geq\rho_n\geq 1 \qquad ;\ (n\geq 1), \tag{2.1}$$

and there is a finite number $a\geq 1$ for which

$$(\rho_{\mu_n}/\rho_{\mu_{n-1}})^{\mu_{n-1}}\leq a^{\nu_n} \qquad ;\ (n\geq 2). \tag{2.2}$$

In view of (1.4), (1.5), (1.6), and (2.1) it can be verified that

$$|p_{m,k}|\leq(k^{\mu_n})\rho_{\mu_n}^{\mu_n-k};\ (0\leq k\leq d_m,\ \mu_{n-1}<m\leq\mu_n;\ n\geq 1). \tag{2.3}$$

Inserting (2.3) in (1.7) we find that

$$\delta_0=1;\ \delta_n<\lambda_n\rho_{\mu_n}^{1/2\nu_n(\nu_n-1)}=\lambda_n T_n\rho_{\mu_n}^{-\nu_n} \qquad ;\ (n\geq 1), \tag{2.4}$$

---

[4] c. f. Ibrahim [2; p. 282].

[5] It should be observed that a substitution $z=k_{z'}+b$ transforms the circles $|z-b|=k$, $|z-b|=kn^{\alpha}$ onto the respective circles $|z'|=1$, $|z'|=n^{\alpha}$! Also, according to Whittaker's theorem [4; p. 12] such substitution does not affect the order of the basic set. Hence there is no loss of generality in assuming the zeros to lie in $|z|\leq 1$ in Theorem 1 and in $|z|\leq n^{\alpha}$ in Theorem 2. We note also that, as far as the order is concerned, the results of Theorems 1 and 2 above reduce to those of Nassif [3; Theorems 6.1, 6.2] when $l=1$.

where

$$\lambda_0=T_0=1,\ \lambda_n=\nu_n!\binom{\mu_n}{1}\binom{\mu_n}{2}\cdots\cdots\binom{\mu_n}{\nu_n-1};\ T_n=\rho_{\mu_n}^{\nu_n(\nu_n+1)/2}\ ;\ (n\geq 1) \tag{2.5}$$

Suppose that z^n admits the representation

$$z^n=\sum_k \pi_{n,k}p_k(z). \tag{2.6}$$

The required lemma is concerned with the coefficients $(\pi_{n,k})$:

LEMMA 1: *Let the set* $\{p_n(z)\}$ *be as in Theorem* 1 *and suppose that the zeros of the polynomial* $p_n(z)$ *lie in* $|z|\leq\rho_n$. *Then the coefficients* $(\pi_{n,k})$ *satisfy the inequality*

$$|\pi_{\mu_n-i,\mu_k-j}|<\frac{i!(\mu_n-i)!\lambda_k\lambda_{k+1}\cdots\lambda_n,\ T_kT_{k+1}\cdots T_n}{\rho_{\mu_n}^i(\mu_k)!\delta_k\delta_{k+1}\cdots\delta_n}\,c^{\mu_n-\mu_k}, \tag{2.7}$$

for $0\leq i\leq\nu_n-1;\ 0\leq j\leq\nu_k-1;\ n\geq k\geq 0$, *where the constant* c *is fixed by*

$$c=\frac{a}{\log\,(1+i/l)}>a\geq 1. \tag{2.8}$$

PROOF: It should be observed, first of all, that the matrix $(\pi_{n,k})$, which, according to (2.6), is the unique inverse of the matrix of coefficients $(p_{n,k})$, is of the same structure as this last matrix. Thus, carrying out the product of the matrices (Δ_k) with their corresponding matrices in $(\pi_{n,k})$, we get $\pi_{0,0}=1$, and

$$(p_{\mu_{k-r},\mu_k-s})(\pi_{\mu_k-t,\mu_k-u})=I_k,\ (0\leq r,s,t,u\leq\nu_k-1;\ k\geq 1),$$

where I_k is the unit matrix of order ν_k. Applying (2.1), (2.3), (2.4), and (2.5) the following inequality is obtained.

$$|\Pi_{\mu_k-i,\mu_k-j}|<\frac{i!(\mu_{k-i})!\lambda_kT_k}{(\mu_k)!\rho_{\mu_k}^i\delta_k};\ (0\leq i,\ j\leq\nu_k-1;\ k\geq 1). \tag{2.9}$$

The product is then carried out with respect the remaining elements of the matrices $(p_{n,k})$ and $(\Pi_{n,k})$. When $k\geq 1$, the following equations are formed.

$$\sum_{s=\mu_{k-1}+1}^{\mu_n} p_{\mu_n-r,s}\Pi_{s,\mu_k-j}=0;\ r=0,\ 1,\ \ldots,\ \nu_n-1;\ (0\leq j\leq\nu_k-1;\ n\geq k\geq 1). \tag{2.10}$$

These equations can be solved for the coefficients $(\Pi_{\mu_n-i,\mu_k-j});\ i=0,\ 1,\ \ldots,\ \nu_n-1$, since $\delta_n>0$. Appealing to the relations (2.1), (2.3), and (2.5) we obtain

$$|\Pi_{\mu_n-i,\mu_k-j}|\leq\frac{\lambda_nT_n}{\binom{\mu_n}{i}\rho_{\mu_n}^{i+\nu_n}\delta_n}\sum_{s=\mu_{k-1}+1}^{\mu_n-1}\binom{\mu_n}{s}\rho_{\mu_n}^{\mu_n-s}|\Pi_{s,\mu_k-j}|, \tag{2.11}$$

for $0\leq i\leq\nu_n-1;\ 0\leq j\leq\nu_k-1;\ n\geq k\geq 1$. The inequality (2.7) of the lemma will be deduced from (2.11) when $k\geq 1$. In fact, it is seen from (2.9) that (2.7) is true for $n=k$. Also, putting $n=k+1$ in (2.11) and applying (1.3), (2.1), (2.2), (2.8), and (2.9) it can be verified that (2.7) is also satisfied for

$n=k+1$. Moreover, suppose that (2.7) is valid for $n=k, k+1, \ldots, m-1$, then by application o (1.2), (1.3), (2.1), (2.2), (2.4), (2.7), (2.8), and (2.11) and by simple calculation we can arrive at th following relation.

$$|\Pi_{\mu_m-i,\mu_k-j}| < \frac{T_m\lambda_m}{\binom{\mu_m}{i}\rho_{\mu_m}^{i+\nu_m}\delta_m}\sum_{n=k}^{m-1}\sum_{r=0}^{\nu_n-1}\binom{\mu_m}{\mu_n-r}\rho_{\mu_m}^{\mu_m-\mu_n+r}|\Pi_{\mu_n-r,\mu_k-j}|$$

$$< \frac{i!(\mu_m-i)!\lambda_k\lambda_{k+1}\ldots\lambda_m\cdot T_kT_{k+1}\ldots T_m}{(\mu_k)!\rho_{\mu_m}^i\,\delta_k\delta_{k+1}\ldots\delta_m}c^{\mu_m-\mu_k}.$$

Hence the inequality (2.7) of the lemma is true for $n\geq k\geq 1$.

Now, when $k=0$, (and hence $j=0$), the eqs (2.10) assume the form

$$\sum_{s=0}^{\mu_n}P_{\mu_n-r,s}\Pi_{s,0}=0;\ (r=0,1,\ldots,\nu_n-1;\ n>0).$$

Solving these equations for the coefficients $(\Pi_{\mu_n-i,0})$; $i=0,1,\ldots\nu_n-1$, and proceeding in the same way as before, we easily obtain the inequality

$$|\pi_{\mu_n-i,0}| < \frac{i!(\mu_n-i)!\lambda_1\lambda_2\ldots\lambda_n\cdot T_1T_2\ldots T_n}{\rho_{\mu_n}^i\,\delta_1\delta_2\ldots\delta_n}c_{\mu_n}, \tag{2.12}$$

for $0\leq i\leq\nu_n-1$; $n>0$. Noting that $\lambda_0=T_0=\lambda_0=1$; $\mu_0=0$ it will be seen that (2.12) is merely the inequality (2.7) for $k=0$. The lemma is therefore established.

3. Proof of Theorem 1

We shall suppose here that the zeros of the polynomials $\{p_n(z)\}$ all lie in $|z|\leq 1$. If $M_n(r)$ denotes the maximum value of $|p_n(z)|$ in $|z|\leq r$; $r>0$, then from (1.5) and (1.6) we have

$$M_{\mu_k-j}(r)\leq(r+1)^{\mu_k}\quad;\ (0\leq j\leq\nu_k-1;\ k\geq 1). \tag{3.1}$$

Moreover, putting[6] $\rho_n=1$ in (2.4), (2.5), and (2.7) we find that

$$\begin{cases}\delta_n<\lambda_n<l\mu_n^{\nu_n(l-1)/2},\\ |\pi_{\mu_n-i,\mu_k-j}|<\dfrac{i!(\mu_n-i)!\lambda_k\lambda_{k+1}\ldots\lambda_n}{(\mu_k)!\delta_k\delta_{k+1}\ldots\delta_n}c^{\mu_n-\mu_k},\\ (0\leq i\leq\nu_n-1;\ 0\leq j\leq\nu_k-1;\ n\geq k\geq 0);\ c=\dfrac{1}{\log(1+1/l)}.\end{cases} \tag{3.2}$$

Therefore

$$\lambda_1\lambda_2\ldots\lambda_n<l^n\mu_n^{\nu_n(l-1)/2}\quad;\ (n\geq 1). \tag{3.3}$$

In the usual notation, the Cannon sum for the set $\{p_n(z)\}$ is

[6] The restrictions (2.1) and (2.2) for the numbers (ρ_n) are satisfied in this case with $a=1$.

(3.4) $$\omega_n(r) = \sum_k |\pi_{n,k}| M_k(r).$$

Introduction of (3.1), (3.2), and (3.3) in (3.4) easily leads to the following relation

$$\omega_{\mu_n - i}(r) < l! \frac{(\mu_n - i)! \mu_n^{\nu_n(l-1)/2}}{} l^n \cdot c^{\mu_n} \exp\left(\frac{r+1}{c}\right);$$

for $0 \leq i \leq \nu_n - 1$; $n \geq 1$. The order ω of the basic set $\{p_n(z)\}$ can be evaluated from this relation by application of (1.3) and (1.8), whereby we obtain

(3.5) $$\omega \leq \frac{1}{2}(l+1) - \frac{\sigma}{l},$$

as required by Theorem 1.

4. Example

To complete the proof of the theorem a basic set is constructed, according to the conditions of the theorem, such that its order $\omega = \frac{1}{2}(l+1) - \frac{\sigma}{l}$. The following lemma is first proved.

LEMMA 2: *When* $l \geq 2$, *the function* $E(z) = \sum_{n=0}^{\infty} z^{nl}/(nl)!$ *has at least one zero inside the circle* $|z| = \{(l+1)!\}^{1/l}$.

PROOF: When $l = 2$, $E(z) = \cosh z$, which obviously has the required property. Therefore we shall assume that $l \geq 3$ and put

(4.1) $$z^l = t;\ E(z) = e(t);\ f(t) = 1 + \frac{t}{l!} + \frac{t^2}{(2l)!};\ e(t) = f(t) + r(t).$$

Now, it is easily seen that when $l \geq 3$, the function $f(t)$ has a zero in $-(l+1)! < t < 0$, and therefore $f(t)$ has at least one zero in the circle $|t| = (l+1)!$. Moreover, by actual calculation it can be verified from (4.1) that

$$|f(t)| > 3/2 \text{ on } |t| = (l+1)! \quad ;\ (l \geq 3),$$

and that

$$\max_{|t| = (l+1)!} |r(t)| < 2/9 \quad ;\ (l \leq 3).$$

Hence, the required result follows by an application of Rouche's theorem.

Now, it is easily seen that the following set $\{p_n(z)\}$ satisfies the conditions of Theorem 1.

(4.2) $$p_0(z) = 1;\ p_{nl-r}(z) = (1 + \epsilon^r z)^{nl};\ (0 \leq r \leq l-1;\ n \geq 1),$$

where $\epsilon = \exp(2i\pi/l)$; $l \geq 2$. It is also clear that the zeros of the polynomials all lie on $|z| = 1$. From (4.2) we have

(4.3) $$p_{nl-r,k} = \binom{nl}{k} \bar{\epsilon}^{rk};\ (0 \leq k \leq nl;\ n \geq 1;\ 0 \leq r \leq l-1).$$

Whence, in the notation (2.4) and (2.5), we obtain

$$\text{(4.4)}\qquad \begin{cases} \delta_n = (l!)^{-1}\lambda_n 2^{l(l-1)/2} \prod_{j=1}^{l-1} |\sin\,(j\pi/l)|^j; \\ \lambda_n = l!\binom{nl}{1}\binom{nl}{2}\cdots\binom{nl}{l-1}. \end{cases}$$

Simple calculation based on (4.4) leads to the fact that $\sigma = \frac{1}{2}l(l-1)$. Hence, in view of (3.5), we have to prove that $\omega = 1$.

In fact, from the matrix product $(p_{n,k})(\pi_{n,k}) = I$ it follows that

$$\text{(4.5)}\qquad \sum_{k=0}^{nl} p_{nl-r,k}\pi_{k,0} = 0,\ (r = 0, 1, \ldots, l-1;\ n \geqslant 1).$$

Inserting (4.3) in (4.5), adding the results corresponding to $r = 0, 1, \ldots, l-1$, and putting

$$\text{(4.6)}\qquad b_k = \frac{\pi_{kl,0}}{(kl)!},\ (k \geqslant 1);\ b_0 = \pi_{0,0} = 1,$$

we are led to the following relations

$$\text{(4.7)}\qquad \sum_{k=0}^{n} \frac{b_{n-k}}{(kl)!} = 0,\ (n \geqslant 1).$$

Writing $E(z) = \sum_{n=0}^{\infty} z^{nl}/(nl)!$ and $G(z) = \sum_{n=0}^{\infty} b_n z^{nl}$, we see that (4.7) implies that $G(z) = 1/E(z)$. Hence, by Lemma 2, we infer that $G(z)$ is regular in $|z| = \rho$, where $\rho < \{(l+1)!\}^{1/l}$. That is to say

$$\text{(4.8)}\qquad \limsup_{n\to\infty} |b_n|^{1/n} = 1/\rho.$$

Finally, from (4.6) and (4.8) we can deduce that $\omega = 1$, and the proof of Theorem 1 is therefore complete.

5. Proof of Theorem 2

We now suppose that the zeros of the polynomial $p_n(z)$, belonging to the regular set $\{p_n(z)\}$, lie in $|z| \leqslant n^\alpha$, where α is a positive number. Hence, in the notation of (3.1), we see that

$$\text{(5.1)}\qquad M_0(r) = 1,\ M_{\mu_k - j}(r) \leqslant \{r + (\mu_k)^\alpha\}^{\mu_k};\ (0 \leqslant j \leqslant \nu_k - 1;\ k \geqslant 1).$$

Moreover, putting[7] $\rho_n = n^\alpha$ in (2.4) and (2.7) we obtain

$$\text{(5.2)}\qquad \begin{cases} \delta_n < \lambda_n \mu_n^{\alpha\nu_n(\nu_n-1)/2} \\ |\pi_{\mu_n - i,\,\mu_k - j}| < \dfrac{i!(\mu_{n-1})!\lambda_k\lambda_{k+1}\ldots\lambda_n,\ U_kU_{k+1}\ldots U_n}{(\mu_k)!\delta_k\delta_{k+1}\ldots\delta_n}\, c^{\mu_n - \mu_k} \\ (0 \leqslant i \leqslant \nu_n - 1;\ 0 \leqslant j \leqslant \nu_k - 1;\ n \geqslant k \geqslant 0), \end{cases}$$

where

[7] The restrictions (2.1) and (2.2) for the numbers (ρ_n) are still here satisfied with $\alpha = \alpha' > 1$.

$$c=\frac{e^{\alpha}}{\log\,(1+i/l)},\ U_k=\mu_k^{\alpha\nu_k(\nu_k+1)/2}\qquad ;\ (k\geqslant 1),$$

so that

$$U_1U_2\ .\ .\ .\ .\ .\ U_n<\mu_n^{\alpha\mu_n(l+1)/2}. \tag{5.3}$$

A combination of (3.3), (3.4), (5.1), (5.2), and (5.3) easily yields

$$\omega_{\mu_n-i}(r)<l!\exp\left\{\frac{e^{\alpha}\,(r+1)}{c}\right\}l_{\mu_nc}^{n\alpha l}\mu_n\frac{(\mu_n-i)!\mu_n^{\mu_n\{\alpha(l+1)+l-1\}/2}}{\delta_1\delta_2\ .\ .\ .\ \delta_n}, \tag{5.4}$$

for $0\leqslant i\leqslant \nu_n-1$; $n\geqslant 1$. The inequality of Theorem 2, namely

$$\omega\leqslant\frac{1}{2}\,(\alpha+1)\,(l+1)-\frac{\sigma}{l},$$

for the order of the basic set $\{p_n(z)\}$, can be easily derived from the relation (5.4). Theorem 2 is therefore established.

6. References

[1] Eweida, M. T., Duke Math J. **14**, 865–875 (1947).
[2] Ibrahim, S. A., J. of the London Math. Soc. **34**, 281–288 (1959).
[3] Nassif, M., Products and zeros of basic sets, Assiut (1958).
[4] Whittaker, J. M., Interpolatory Function Theory (Cambridge Press, 1935).
[5] Whittaker, J. M., Sur les Séries de Base de Polynomes Quelconques, Paris (1949)

(Paper 76B3&4–368)

JOURNAL OF RESEARCH of the National Bureau of Standards—Mathematical Sciences
Vol. 76B, Nos. 3 and 4, July–December 1972

The Stieltjes Constants*

J. J. Y. Liang** and John Todd***

(July 31, 1972)

The Stieltjes Constants are the coefficients in the Laurent expansion of the Riemann Zeta function $\xi(z)$ about its simple pole at $z=1$. They can be represented as the limit of the difference between the sum of the first n terms of a series and the integral of its n-th term.

The first 20 coefficients have been computed to 15 D using the Euler-Maclaurin method. As a by-product the sums of the series

$$\tau_n=\sum_2^\infty (-1)^k(\log k)^n/k$$

have been obtained to 15 D for $n=1(1)20$.

Key words: Bernoulli numbers; Euler-Maclaurin method; Euler transform; Euler's constant; multiple precision package; Riemann zeta function.

1. Introduction.

The Riemann ζ-function defined by

$$\zeta(z)=\sum_0^\infty n^{-z} \tag{1.0}$$

is regular if $x=\mathscr{R}z>1$; a continuation into the half plane $\mathscr{R}z>0$ is given by

$$(2^{1-z}-1)\zeta(z)=\sum_1^\infty (-1)^n n^{-z}. \tag{1.0'}$$

We can avoid the introduction of the idea of continuation by defining $\zeta(z)$ in $\mathscr{R}z>0$, except at $z=1$, by (1.0'). Uniform convergence in $\mathscr{R}z>0$ follows since $\Sigma(-1)^n n^{-x}$ is convergent as an alternating series for $x>0$. (Cf. Knopp. [10, p. 441][1].) The fact that

$$(z-1)\zeta(z)\to 1$$

as $z\to 1$ then follows since $\sum_1^\infty (-1)^n n^{-1}=-\log 2$ and since the exponential series gives

$$\begin{aligned}(2^{1-z}-1)&=e^{(1-z)\log 2}-1\\&=-(z-1)\log 2\left[1+\frac{(z-1)\log 2}{2!}+\ldots\right].\end{aligned}$$

This function has a Laurent expansion about the simple pole at $z=1$:

$$\zeta(z)=\frac{1}{z-1}+A_0+A_1(z-1)+A_2(z-1)^2+A_3(z-1)^3+\ldots \tag{1.1}$$

AMS Subject classification: Primary 10H05, Secondary 65B10, 65B15.
*An invited paper, the preparation of which was supported in part by the National Science Foundation.
**University of South Florida, Tampa, Florida 33620.
***California Institute of Technology, Pasadena, California 91109.
[1] Figures in brackets indicate the literature references at the end of this paper.

where the coefficients are given by

$$A_n = (-1)^n \gamma_n / n!$$

where

$$\gamma_n = \lim_{m \to \infty} \left\{ \sum_{k=1}^{m} \frac{(\log k)^n}{k} - \frac{(\log m)^{n+1}}{n+1} \right\}, \quad n = 0, 1, 2, \ldots . \tag{1.2}$$

(We have changed the notation of Stieltjes and Jensen to agree with that of Hardy and Briggs and Chowla.)

It is clear that $\gamma_0 = \gamma$, Euler's constant. Stieltjes proposed [3, I, Letter 77] the computation of (the first five of) these constants and gave [3, I, Letter 71] the value

$$-A_1 = \gamma_1 = -0.07281\ 5520$$

and estimated

$$A_2 \doteqdot -0.0047$$

Hermite [3, I, Letter 74] wrote that, during a session of the Academy of Sciences, he found some objections to Stieltjes' proof of (1.2) and that he had obtained a more correct proof. Stieltjes [3, I, Letter 75] gave a detailed proof of (1.2). This result was also obtained by Jensen ([9] and [3, II, p. 451]) and stated by Hardy [8] and by Ramanujan [12]. Two proofs have been given recently by Briggs and Chowla [2]—we reproduce one of these in section 4.

We note that tables of the A's have been prepared by Jensen [9], who gives A_i, $i = 1\ (1)\ 9$ to $9D$, and by Gram [7], who gives $A_i, i = 1(1)\ 16$ to $16D$. In section 9 we describe Gram's method of computation. In each table the later coefficients are zero to the number of decimals given. We shall see that these results appear to be correct. It is interesting to note that one of the reasons for calculating these constants was the determination of the small complex zeros of $\zeta(s)$. (See also Lammel [21].) The apparently irregular behavior of these coefficients has been investigated analytically by Briggs [1].

Our first calculations were based on extensions by Hardy [8] of a result of Vacca [15] which leads to expressions for the γ's in terms of elementary constants and the series

$$\tau_n = \sum_{k=2}^{\infty} (-1)^k \frac{(\log k)^n}{k}, \qquad n = 0,1,2, \ldots .$$

For instance, Hardy gave

$$\tau_0 = 1 - \log 2 = 1 - 0.69314\ 718 = 0.30685\ 282$$

$$\tau_1 = -\frac{1}{2}(\log 2)^2 + \gamma \log 2 = 0.15986\ 890$$

$$\tau_2 = -\frac{1}{3}(\log 2)^3 + \gamma\ (\log 2)^2 + 2\gamma_1 \log 2 = 0.06537\ 259.$$

(We have corrected errors of sign in Hardy's expressions for τ_1, τ_2.) The general form of the relations connecting the τ's and γ's has been given by Briggs and Chowla and is discussed in section 3. We calculated the τ's by use of a delayed Euler transform and then obtained the γ's by solving the triangular system of linear equations. However (see sec. 5) this method proved unsatisfactory and we made direct calculations of the γ's using the Euler-Maclaurin series. This is a natural extension of the calculations made by Knuth [11] in which he obtained γ to 1271 D. Error estimates

are easier to obtain in the Euler-Maclaurin computation, but these computations are more lengthy than the direct ones.

We have only given error estimates for the cases of γ_1, γ_2 but representative computations for the later γ's suggest that our values of γ are secure. We have therefore to regard the γ's as our primary results and the τ's as derived from them.

Our interest in the present problem was aroused by an entry (giving γ_1) in the tables of Wheelon [6] which we used for exercises in courses dealing with the Euler transforms.

Professor H. Zassenhaus has called our attention to rapidly convergent series for γ obtained by Jacobstahl [24] (see also Addison [25]) which might lead to alternative attacks on the higher γ's.

2. The γ's

The limits by which the γ's are defined clearly exist, e.g., in view of a supplement to the integral test (Knopp, [10, p. 295]), since

$$f_k(x) = \frac{(\log x)^k}{x}$$

is ultimately decreasing because

$$f_k'(x) = \frac{(\log x)^{k-1}}{x^2}[k - \log x],$$

which is negative for $x > e^k$. The sequences $\gamma_k^{(1)}, \gamma_k^{(2)}, \gamma_k^{(3)}, \ldots$ where

$$\gamma_k^{(n)} = \sum_{\nu=1}^{n} \frac{(\log \nu)^k}{\nu} - \frac{1}{k+1}(\log n)^{k+1} - \gamma_k$$

are ultimately monotone decreasing since

$$\begin{aligned}\gamma_k^{(n)} - \gamma_k^{(n+1)} &= \frac{1}{k+1}(\log (n+1))^{k+1} - \frac{1}{k+1}(\log n)^{k+1} - \frac{(\log (n+1))^k}{n+1} \\ &= \int_n^{n+1} f_k(t)\,dt - f_k(n+1) \qquad (2.1) \\ &> 0 \text{ if } f_k(t) \text{ is decreasing for } t \geq n, \text{ i.e., if } n \geq e^k.\end{aligned}$$

We shall now investigate the size of $\gamma_k^{(n)}$, i.e., the speed of convergence of

$$\sum_{\nu=1}^{n} \frac{(\log \nu)^k}{\nu} - \frac{1}{k+1}(\log n)^{k+1}$$

to its limit γ_k. It will appear that

$$\gamma_k^{(n)} = \frac{1}{2}\frac{(\log n)^k}{n} + 0(n^{-2+\epsilon}) \qquad (2.2)$$

which means that direct computation of γ_k does not come in question. We follow the analysis given in the case $k = 0$ by Francis and Littlewood [5, (Question B7, p. 2, 19)].

We require the following lemma which is a simple case of the Euler-Maclaurin formula. (See Hardy [22, p. 300, Ex. 4]).

479-206 O - 73 - 5

LEMMA. *If* g *is three times differentiable in* $[0, 1]$ *then*

$$g(1) = g(0) + \frac{1}{2}[g'(0) + g'(1)] - \frac{1}{12}g'''(\theta)$$

where $0 < \theta < 1$.

PROOF. Consider

$$G(x) = g(x) - g(0) - \frac{1}{2}x[g'(0) + g'(x)] - x^3\left[g(1) - g(0) - \frac{1}{2}\{g'(0) + g'(1)\}\right].$$

We have

$$G'(x) = g'(x) - \frac{1}{2}[g'(0) + g'(x)] - \frac{1}{2}xg''(x) - 3x^2\left[g(1) - g(0) - \frac{1}{2}\{g'(0) + g'(1)\}\right]$$

and

$$G''(x) = -\frac{1}{2}xg'''(x) - 6x\left[g(1) - g(0) - \frac{1}{2}\{g'(0) + g'(1)\}\right].$$

We note that $G(0) = G(1) = 0$ and so $G'(\theta_0) = 0$ for some θ_0, $0 < \theta_0 < 1$. Also $G'(0) = 0$. Hence $G''(\theta) = 0$ for some θ, $0 < \theta < \theta_1$. Since $\theta \neq 0$ we have

$$g'''(\theta) = -12\left[g(1) - g(0) - \frac{1}{2}\{g'(0) + g'(1)\}\right]$$

which is the result required.

Applying the lemma in the case

$$g(x) \equiv (\log(x+n))^{k+1}/(k+1),$$

gives (cf. (2.1)):

$$\frac{(\log(n+1))^{k+1}}{k+1} - \frac{(\log n)^{k+1}}{k+1} - \frac{(\log(n+1))^k}{n+1} = \gamma_k^{(n)} - \gamma_k^{(n+1)} \tag{2.3}$$

$$= \frac{1}{2}[f_k(n) - f_k(n+1)] + e(n,k)$$

where

$$e(n,k) = O((\log n)^k/n^3).$$

Summing (2.3) for $n = m$ to $m = M-1$ gives for a suitable K,

$$\gamma_k^{(m)} - \gamma_k^{(M)} \geqslant \frac{1}{2}[f_k(m) - f_k(M)] - K\sum_{n=m}^{M-1}(\log n)^k/n^3$$

and

$$\gamma_k^{(m)} - \gamma_k^{(M)} \leqslant \frac{1}{2}[f_k(m) - f_k(M)] + K\sum_{n=m}^{M-1}(\log n)^k/n^3.$$

Letting $M \to \infty$ gives, for any $\epsilon > 0$,

$$\gamma_k^{(m)} = \frac{1}{2} f_k(m) + O(m^{-2+\epsilon})$$

which is the result (2.2) announced.

3. Derivation of the τ–γ Relations

We have seen that as $n \to \infty$:

$$\sum_{\nu=1}^{n} \frac{(\log \nu)^k}{\nu} = \frac{(\log n)^{k+1}}{k+1} + \gamma_k + o(1). \tag{3.1}$$

Hence

$$\sum_{\nu=1}^{2n} \frac{(\log \nu)^k}{\nu} = \frac{(\log 2n)^{k+1}}{k+1} + \gamma_k + o(1). \tag{3.2}$$

The binomial series for $(\log 2\nu)^k = (\log 2 + \log \nu)^k$ gives

$$\begin{aligned} 2\sum_{\nu=1}^{n} \frac{(\log 2\nu)^k}{2\nu} &= \sum_{\nu=1}^{n} \frac{1}{\nu} \left\{ \sum_{t=0}^{k} \binom{k}{t} (\log 2)^{k-t} (\log \nu)^t \right\} \\ &= \sum_{t=0}^{k} \binom{k}{t} (\log 2)^{k-t} \sum_{\nu=1}^{n} \frac{(\log \nu)^t}{\nu} \\ &= \sum_{t=0}^{k} \binom{k}{t} (\log 2)^{k-t} \left[\frac{(\log n)^{t+1}}{t+1} + \gamma_t + o(1) \right] \end{aligned} \tag{3.3}$$

where we have used (3.1) in the last stage.

Integrating (with respect to β) the binomial expansion of $(a+\beta)^k$ from 0 to b, or otherwise, we find

$$\sum_{t=0}^{k} \binom{k}{t} a^{k-t} \frac{b^{t+1}}{t+1} = \left[\frac{(a+\beta)^{k+1}}{k+1} \right]_0^b = \frac{(a+b)^{k+1}}{k+1} - \frac{a^{k+1}}{k+1}.$$

We use this with

$$a = \log 2, \qquad b = \log n$$

to get

$$\sum_{t=0}^{k} \binom{k}{t} (\log 2)^{k-t} \frac{(\log n)^{t+1}}{t+1} = \frac{(\log 2n)^{k+1}}{k+1} - \frac{(\log 2)^{k+1}}{k+1}. \tag{3.4}$$

Substituting from (3.4) in (3.3) we get

$$2\sum_{\nu=1}^{n} \frac{(\log 2\nu)^k}{2\nu} = \frac{(\log 2n)^{k+1}}{k+1} - \frac{(\log 2)^{k+1}}{k+1} + \sum_{t=0}^{k} \binom{k}{t} (\log 2)^{k-t} \gamma_t + o(1) \tag{3.5}$$

and then subtracting (3.5) from (3.2) gives

$$-\sum_{\nu=1}^{2n} \frac{(-1)^{\nu-1} (\log \nu)^k}{\nu} = -\frac{(\log 2)^{k+1}}{k+1} + \sum_{t=0}^{k-1} \binom{k}{t} (\log 2)^{k-t} \gamma_t + o(1),$$

the last term ($t=k$) in the summation with respect to t cancelling the γ_k on the right of (3.2). We now let $n\to\infty$ and find

$$\tau_k=-\frac{(\log 2)^{k+1}}{k+1}+\sum_{t=0}^{k-1}\binom{k}{t}(\log 2)^{k-t}\gamma_t. \tag{3.6}$$

For $k=1, 2, \ldots$ we find

$$\tau_1=-\frac{1}{2}(\log 2)^2+\gamma\log 2$$

$$\tau_2=-\frac{1}{3}(\log 2)^3+\gamma(\log 2)^2+2\gamma_1\log 2$$

$$\tau_3=-\frac{1}{4}(\log 2)^4+\gamma(\log 2)^3+3\gamma_1(\log 2)^2+3\gamma_2\log 2.$$

The result for $k=0$ is, trivially,

$$\tau_0=\left(\frac{1}{2}-\frac{1}{3}+\frac{1}{4}\ldots\right)=1-\left(1-\frac{1}{2}+\frac{1}{3}-\ldots\right)=-\log 2+1.$$

4. Proof of Representation

If we differentiate the relation

$$g(z)\equiv(2^{1-z}-1)\zeta(s)=\sum_{n=1}^{\infty}(-1)^n n^{-z}$$

k times with respect to z we get

$$g^{(k)}(z)=(-1)^k\sum_{n=1}^{\infty}(-1)^n(\log n)^k/n^z$$

and, in particular,

$$g^{(k)}(1)=(-1)^k\tau_k. \tag{4.1}$$

We have

$$(2^{1-z}-1)=e^{-(z-1)\log 2}-1=\sum_{n=1}^{\infty}\frac{(-1)^n(\log 2)^n}{n!}(z-1)^n$$

and multiplying this by

$$\zeta(s)=(z-1)^{-1}+\sum_{0}^{\infty}A_n(z-1)^n$$

we get

$$g(z)=\sum_{k=0}^{\infty}\sum_{t=1}^{k+1}\frac{(-1)^t(\log 2)^t}{t!}A_{k-t}(z-1)^k$$

where $A_{-1}=1$. Hence

$$g^{(k)}(1)=k!\sum_{t=1}^{k+1}\frac{(-1)^t(\log 2)^t}{t!}A_{k-t}. \tag{4.2}$$

We now use the representation (3.6) for τ_k to find from (4.1) and (4.2)

$$(-1)^k\left[-\frac{(\log 2)^{k+1}}{k+1}+\sum_{t=0}^{k-1}\binom{k}{t}(\log 2)^{k-t}\gamma_t\right]=k!\sum_{t=1}^{k+1}\frac{(-1)^t(\log 2)^t}{t!}A_{k-t}. \tag{4.3}$$

If we use this relation for $k=1, 2, 3$ respectively we find

$$\gamma=\gamma_0=A_0,$$

$$\gamma_1=-A_1,$$

$$\gamma_2=2!A_2.$$

These suggest the general solution

$$\gamma_n=(-1)^n n!\,A_n$$

which is readily verified. The term outside the summation on the left in (4.3),

$$(-1)^{k+1}\,(\log 2)^{k+1}\,/(k+1),$$

is the same as the term for $t=k+1$ in the summation as the right. Then the t-th term on the left, for $t=0, 1, \ldots, k-1$, can be identified with the $(k-t)$-th term on the right.

5. Calculation of the τ's by Euler Transforms

Knowing the efficiency of the Euler transform in handling the logarithmic series, it is natural to apply this to calculate the τ's.

Specifically, the Euler transform of $\Sigma(-1)^n/n$, when there is a delay of r, is

$$1-\frac{1}{2}+\frac{1}{3}-\ldots+(-1)^{r-1}\frac{1}{r}+(-1)^r\frac{1}{r+1}\left[\frac{1}{2}+\frac{1}{4}\cdot\frac{1}{r+2}+\frac{1}{8}\cdot\frac{1\cdot 2}{r+2\cdot r+3}+\ldots\right].$$

Since the sequence $\{n^{-1}\}$ is completely monotone, all the terms in the tail of the transform are of one sign and we can readily estimate the error. When we take n terms in the tail this error is

$$\log 2-(s(r)+c(r,n))=(-1)^r\frac{1}{r+1}\left[\frac{1}{2^{n+1}}\cdot\frac{1\cdot 2\cdot\cdots\cdot n}{r+2\cdot r+3\cdot\cdots\cdot(r+n+1)}+\ldots\right]$$

and is less in absolute value than

$$((r+1)2^{n+1})^{-1}.$$

Various hypotheses lead to various optimal choices of r,n. But these are not critical. It has also been pointed out that the optimal choice of r,n did not seem to depend critically on the (alternating) series being summed (see, e.g. Todd, [14]).

The Euler transform is, of course, convergent if the original series is, but improvement in the speed of convergence of $\Sigma(-1)^n u_n$ has only been established in case the sequence $\{u_n\}$ is completely monotone and if $(u_{n+1}/u_n)\geqslant k>\frac{1}{2}$ (see, e.g., Knopp [10, p. 253]).

When u_n is completely monotone the terms in the remainder are all of one sign and error estimates are easy to obtain. However the terms in τ_k are not even monotone: since

$$\frac{d}{dx}\left[\frac{(\log x)^k}{x}\right]=\frac{(\log x)^{k-1}}{x^2}[k-\log x]$$

we expect the terms in τ_k to increase up to about the $[e^k]$ -th and then to decrease. Continuing we see that

$$\frac{d^2}{dx^2}\left[\frac{(\log x)^k}{x}\right]=\frac{(\log x)^{k-2}}{x^3}[2(\log x)^2-3k\log x+k(k-1)]$$

which vanishes for

$$\log x=\frac{3}{4}k\pm\frac{1}{4}\sqrt{k^2+8}$$

so that the second difference may be expected to decrease, become negative and then increase to positive values.

We have not obtained useful error bounds for the Euler transforms of the $\tau_k(k>1)$.

We shall discuss these matters in some detail elsewhere.

6. Calculation of the γ's from the τ's

In section 3 we have derived the relations:

$$\tau_n=-(\log 2)^{n+1}/(n+1)+\sum_{r=0}^{n-1}\binom{n}{r}(\log 2)^{n-r}\gamma_r,\ n=1,2,3,\ \ldots, \tag{3.6}$$

which we have used to compute the τ's from the γ's. From this triangular system of linear equations it is also trivial to obtain the γ's, given the τ's. Although the solution of a triangular system is a numerically stable process (see, e.g., Wilkinson [17, 247]) we have obtained the exact inverse of the 20×20 matrix

$$A=[a_{ij}],\ a_{ij}=\begin{cases}\binom{i}{j} & j=0,1,2,\ \ldots,i-1\\ 0 & j>i\end{cases}\qquad i=1,2,\ \ldots$$

so that the γ's can be obtained with less rounding from the τ's.

$$\begin{bmatrix}\gamma_0\\ \gamma_1/\log 2\\ \ldots\\ \gamma_{n-1}/(\log 2)^{n-1}\end{bmatrix}=A^{-1}\times\begin{bmatrix}\frac{1}{2}\log 2+\tau_1/\log 2\\ \frac{1}{3}\log 2+\tau_2/(\log 2)^2\\ \ldots\\ \frac{1}{n+1}\log 2+\tau_n/(\log 2)^n\end{bmatrix}$$

It is easy to find A^{-1} when $n=10$, e.g., by inverting A and identifying the appropriate rational entries in A^{-1}. Specifically, we find

$$A^{-1}=\begin{bmatrix}
1 & & & & & & & & & \\
-\frac{1}{2} & \frac{1}{2} & & & & & & & & \\
\frac{1}{6} & -\frac{1}{2} & \frac{1}{3} & & & & & & & \\
0 & \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} & & & & & & \\
-\frac{1}{30} & 0 & \frac{1}{3} & -\frac{1}{2} & \frac{1}{5} & & & & & \\
0 & -\frac{1}{12} & 0 & \frac{5}{12} & -\frac{1}{2} & \frac{1}{6} & & & & \\
\frac{1}{42} & 0 & -\frac{1}{6} & 0 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{7} & & & \\
0 & \frac{1}{12} & 0 & -\frac{7}{24} & 0 & \frac{7}{12} & -\frac{1}{2} & \frac{1}{8} & & \\
-\frac{1}{30} & 0 & \frac{2}{9} & 0 & -\frac{7}{15} & 0 & \frac{2}{3} & -\frac{1}{2} & \frac{1}{9} & \\
0 & -\frac{3}{20} & 0 & \frac{1}{2} & 0 & -\frac{7}{10} & 0 & \frac{3}{4} & -\frac{1}{2} & \frac{1}{10}
\end{bmatrix}$$

This naive approach will not work to find A^{-1} in the 20×20 case, owing to the condition of A. We are indebted to Morris Newman for providing us with the inverse in the general case:

$$[A^{-1}]_{ij}=\begin{cases} 0 \text{ if } i<j \\ \binom{i}{j}\frac{B_{i-j}}{i}, i\geq j \end{cases} \qquad i=1, 2, \ldots$$

where the B_k's are the Bernoulli numbers ($B_0=1$, $B_1=-\frac{1}{2}$, $B_2=\frac{1}{6}$, $B_3=0$, etc.)which are readily available as rational fractions or as decimals. Incidentally the characteristic vectors of A can be expressed as Stirling numbers of the second kind.

7. Calculation of the γ's by Euler-Maclaurin

The Euler-Maclaurin summation formula can be written as

$$\sum_{j=p}^{q} f(j)-\int_p^q f(t)dt=\frac{1}{2}\{f(p)+f(q)\}+\sum_{j=1}^{k}\frac{B_{2j}}{(2j)!}\{f^{(2j-1)}(q)-f^{(2j-1)}(p)\}+R_k(p, q), \qquad (7.0)$$

where the B_k's are the Bernoulli numbers, $B_2=\frac{1}{6}$, $B_4=\frac{-1}{30}$, $B_6=\frac{1}{42}$, ... We give expressions for the remainder:

$$R_k(p, q)=\frac{1}{(2k+1)!}\int_p^q P_{2k+1}(t)\, f^{(2k+1)}(t)\, dt, \qquad (7.1a)$$

where

$$P_{2k+1}(t)=\sum_{\nu=1}^{\infty}\frac{2\sin 2\nu\pi t}{(2\pi\nu)^{2k+1}};$$

$$R_k(p,q)=\frac{-1}{(2k)!}\int_0^1 B_{2k}(t)\left[\sum_{\nu=p}^{q-1}f^{(2k)}(\nu+t)\right]dt, \tag{7.1b}$$

where $B_{2k}(t)$ is the Bernoulli polynomial. For up-to-date discussions of these results see Ostrowski [19].

We use (7.0) in the following way. We first take $p=1$, $q=n$:

$$\mathscr{E}(n)\equiv\sum_{j=1}^{n}f(j)-\int_1^n f(t)dt=\frac{1}{2}[f(n)+f(1)]+\sum_{j=1}^{k}\frac{B_{2j}}{(2j)!}\{f^{(2j-1)}(n)-f^{(2j-1)}(1)\}+R_k(1,n). \tag{7.2a}$$

We next assume f is such that $f^{(\mu)}(t)\to 0$ for all relevant μ and that $R_k(1,\infty)$ exists. Letting $n\to\infty$ gives

$$\lim\left[\sum_{j=1}^{n}f(j)-\int_1^n f(t)dt\right]=\frac{1}{2}f(1)+\sum_{j=1}^{k}\frac{B_{2j}}{(2j)!}\{-f^{(2j-1)}(1)\}+R_k(1,\infty). \tag{7.2b}$$

If we now subtract (7.2a) from (7.2b) we get

$$\lim_{m\to\infty}\mathscr{E}(m)=\mathscr{E}_n-\frac{1}{2}f(n)-\sum_{j=1}^{k}\frac{B_{2j}}{(2j)!}f^{(2j-1)}(n)+R_k(n,\infty). \tag{7.3}$$

This is the formula which is used to calculate the γ's. We have to choose n, k so that $R_k(n,\infty)$ is appropriately small. In the cases with which we are concerned, the series

$$\sum t_j\equiv\sum_{j=1}^{\infty}\frac{B_{2j}}{(2j)!}f^{(2j-1)}(n)$$

is easily seen to be divergent using the well known (Knopp, [10, p. 237]) estimates for B_{2j+2}/B_{2j} and the form of $f^{(2j-1)}(t)$ when $f(t)=(\log t)^k/t$. Indeed, using the expression for $f^{(r)}(t)$ given below (7.11), we find

$$\left|\frac{t_{j+1}}{t_j}\right|=\left|\frac{B_{2j+2}}{B_{2j}}\right|\times\frac{1}{(2j+1)(2j+2)}\times\frac{f^{(2j+1)}(n)}{f^{(2j-1)}(n)}$$

$$\approx\frac{(2j+1)(2j+2)}{4\pi^2}\times\frac{1}{(2j+1)(2j+2)}\times\frac{(2j+1)(2j+2)}{n^2}$$

which tends to infinity with j, n and k being fixed.

We shall discuss the behavior of $R_k(n,\infty)$ in the case of γ_1, γ_2 using Ostrowski's estimate for (7.1b). We shall begin with an examination of γ using (7.1a) which is a convenient way of finding an estimate which we require.

Taking $f(x)=x^{-1}$ in (7.3) we have, with the remainder in the form (7.1a),

$$\gamma=1+\frac{1}{2}+\cdots+\frac{1}{n}-\log n-\frac{1}{2n}+\int_n^{\infty}P_1(t)\left[\frac{-1}{t^2}\right]dt$$

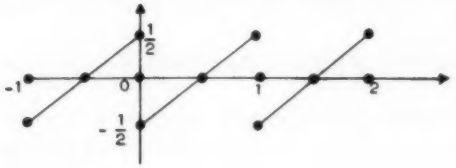
so that

$$1+\frac{1}{2}+\cdots+\frac{1}{n}=\log n+\gamma+\frac{1}{2n}+\int_n^\infty P_1(t)t^{-2}dt.$$

By the Second Mean Value Theorem, there is an $\eta \geq n$ such that

$$\int_n^\infty P_1(t)t^{-2}dt=n^{-2}\int_n^\eta P_1(t)dt.$$

Now it is known that the graph of $P_1(t)$ has the following form:



It follows that $0 \geq \int_n^\eta P_1(t)dt \geq -1/8$. Hence

$$\log n+\gamma+\frac{1}{2n} \geq 1+\frac{1}{2}+\cdots+\frac{1}{n} \geq \log n+\gamma+\frac{1}{2n}-\frac{1}{8n^2}. \tag{7.4}$$

We shall use this result presently.

By a similar argument, starting from

$$\gamma_1=\frac{\log 2}{2}+\frac{\log 3}{3}+\cdots+\frac{\log n}{n}-\frac{1}{2}(\log n)^2-\frac{\log n}{2n}+\int_n^\infty P_1(t)\frac{[1-\log t]}{t^2}dt$$

we obtain

$$\frac{1}{2}(\log n)^2+\gamma_1+\frac{\log n}{2n} \geq \frac{\log 1}{1}+\frac{\log 2}{2}+\ldots+\frac{\log n}{n}$$

$$\geq \frac{1}{2}(\log n)^2+\gamma_1+\frac{\log n}{2n}-\frac{1}{8}\left[\frac{1-\log n}{n^2}\right]. \tag{7.5}$$

(Cf. Boas and Wrench [20].)

We now want to discuss the error in (7.3) in the case of γ_1, i.e., when $f(t) \equiv f_1(t) = (\log t)/t$. We have to find the derivatives of $f_1(t)$.

LEMMA 1. *If* $f_1(x)=(\log x)/x$ *then for* $r \geq 0$ *we have*

$$f_1^{(r)}(x)=\frac{(-1)^r(r!)}{x^{r+1}}[\log x-d_r] \tag{7.6}$$

where $d_0=1$ *and for* $r>0$

$$d_r=1+\frac{1}{2}+\ldots+\frac{1}{r}.$$

PROOF. Induction.

The error estimate given by Ostrowski is

$$|R_k(n,\infty)| \leq \frac{1}{2}\left(1-\frac{1}{2^{2k}}\right)\frac{|B_{2k}|}{(2k)!}|f^{(2k)}(n)| \tag{7.7}$$

and this is valid if $f^{(2k)}(x)$ is monotonic in (n,∞). In our case we shall examine when $f^{(2k+1)}(x)$ is of one sign in (n,∞). In view of Lemma 1 we have to determine a value of x for which

$$\log x > d_{2k+1}.$$

Using the estimate (7.4) we see that we require

$$\log x > \log(2k+1)+\gamma+\frac{1}{2(2k+1)}-\frac{1}{8(2k+1)^2}$$

i.e.,

$$x > (2k+1)\,e^{\gamma}\exp\left(\frac{1}{2(2k+1)}-\frac{1}{8(2k+1)^2}\right).$$

For $k=10$ this gives

$$x > 21\times 1.7811\times 1.025 = 38.34.$$

A rough estimate of the error in the case $k=10$, $n=40$ is, from (7.7),

$$\frac{1}{2}\cdot 1\cdot\frac{|B_{20}|}{(20)!}\cdot\frac{(20)!}{40^{21}}\left[\log 40-\left(1+\frac{1}{2}+\ldots+\frac{1}{20}\right)\right]=\frac{1}{2}\cdot\frac{5.2912\times 10^2}{40^{21}}[3.6889-3.5977]$$

which is certainly negligible to the precision tabulated.

We now turn to the case of γ_2. The analogue of Lemma 1 is

LEMMA 2. *If* $f_2(x)=(\log x)^2/x$ *then for* $r\geq 0$

$$f_2^{(r)}(r)=\frac{(-1)^r r!}{x^{r+1}}[a_r-b_r\log x+(\log x)^2] \tag{7.8}$$

where $a_0=b_0=0$ *and for* $r\geq 0$

$$b_{r+1}=b_r+2/(r+1)\ \textit{giving}\ b_r=2\left[1+\frac{1}{2}+\ldots+\frac{1}{r}\right],\ a_{r+1}=a_r+b_r/(r+1). \tag{7.9}$$

PROOF. Induction.

In order to apply our estimates we have to find a value of $x_0(r)$ such that

$$E(r,x)\equiv[a_r-b_r\log x+(\log x)^2]\geq 0$$

if $x\geq x_0(r)$. From the recurrence relations defining the $\{a_r\}$, $\{b_r\}$ we conclude that $b_r\sim 2\log r$, $a_r\sim(\log r)^2$ which suggests that the critical value will be $O(r)$. Computations also suggest that we may take $x_0=O(r)$. We examine this theoretically.

From the recurrence relations (7.9) we have

$$E(r+1,x)-E(r,x)=\frac{b_r}{r+1}-\frac{2\log x}{r+1}.$$

Summing this we find

$$E(r, x) = (\log x)^2 - 2(\log x)d_r + \sum_{s=1}^{r-1} (b_s/(s+1)).$$

We shall use our estimates (7.4), (7.5) to approximate $E(r, x)$. We have, with errors of O(1),

$$E(r, x) \doteqdot (\log x)^2 - 2(\log x)\{\log r + \gamma\} + 2\sum_{s=1}^{r-1}\left\{\frac{\log(s+1)}{s+1} + \frac{\gamma}{s+1} - \frac{1}{2(s+1)^2}\right\}$$

$$\doteqdot (\log x)^2 - 2(\log x)\{\log r + \gamma\} + (\log r)^2 + 2\gamma \log r.$$

Trying $x = 2r$, i.e., $\log x = \log r + \log 2$, in this we get, to the same accuracy,

$$\begin{aligned} E(r, 2r) &\doteqdot (\log r)^2 + 2(\log r)(\log 2) - 2(\log r)^2 - 2(\log r)(\log 2 + \gamma) \\ &\quad + (\log r)^2 + 2\gamma(\log r) \\ &= 0. \end{aligned}$$

More precise investigations can be made but this discussion seems adequate in the context of practical computation.

We shall now indicate an alternative approach. Let t_r be the larger root of the quadratic $a_r - b_r t + t^2 = 0$, i.e., $t_r = \frac{1}{2}(b_r + \sqrt{b_r^2 - 4a_r})$. Then any number larger than $\exp t_r$ can be taken as x_0. From the recurrence relations (7.9) we have, if $\Delta_r = b_r^2 - 4a_r$,

$$\Delta_{r+1} - \Delta_r = 4(r+1)^{-2}.$$

Summing this we find

$$\Delta_r = 4\left(1 + \frac{1}{2^2} + \cdots + \frac{1}{r^2}\right)$$

$$= 4\left\{\frac{\pi^2}{6} - \left(\frac{1}{(r+1)^2} + \frac{1}{(r+2)^2} + \cdots\right)\right\}$$

$$< \frac{2\pi^2}{3} - 4\left\{\frac{1}{(r+1)(r+2)} + \frac{1}{(r+2)(r+3)} + \cdots\right\}$$

$$= \frac{2\pi^2}{3} - \frac{4}{r+1}.$$

From (7.4) we find

$$t_r < \log r + \gamma + \frac{1}{2r} + \left(\frac{\pi^2}{6} - \frac{1}{r+1}\right)^{1/2}$$

$$< \log r + \gamma + \frac{1}{2r} + \frac{\pi}{\sqrt{6}}\left(1 - \frac{3}{\pi^2(r+1)}\right)$$

$$< \log r + \left(\gamma + \frac{\pi}{\sqrt{6}}\right) + \frac{1}{2r} - \frac{\sqrt{3}}{\sqrt{2}\,\pi(r+1)}$$

$$\doteqdot \log r + 1.8599.$$

This estimate is in agreement with computations. We find

$$a_{25}=0.1296\times 10^{2},\qquad b_{25}=0.7632\times 10^{1},\qquad \log 25=0.3219\times 10^{1},\qquad t_{25}=0.5083\times 10^{1}$$

Hence we can take $x_0(r)=6.5\times r$.

We give here a typical exploratory calculation which we made in the case of γ_2:

$n=400$:	3 Bernoulli terms give	-0.12455	57727	57476	(-4)
which with the head of series gives	$\gamma_2=-0.96903$	63192	87234	(-2)

We use the Ostrowski error estimate (7.7) which in the present case is:

$$\frac{1}{2}\left(1-\frac{1}{2^6}\right)\frac{B_6}{6!}\left[\frac{6!}{n^7}\{a_6-b_6\log n+(\log n)^2\}\right];$$

This is valid if $\log n$ exceeds the larger root of $a_6-b_6\,t+t^2=0$, where $a_6=203/45$ and $b_6=49/10$, which is $t_6=3.6712$. With $n=400$, $\log n \doteq 6$. The term in braces $\{\ldots\}$ is about 11 and so the truncation error estimate is about 10^{-19}—so that the round off errors will dominate. Note that in dealing with the head of the series we are summing 400 terms, each involving the square (or cube) of a logarithm.

We have not made any detailed error analyses for the cases of γ_m, for $m>2$. In our final computations we have taken $n=400$. For instance, in the case of γ_{15},

	3 Bernoulli terms give	0.36042	49967	32360	$(+6)$
which with the head of the series gives	$\gamma_{15}=-0.28369$	16022	44191	(-3)

so that there is a massive cancellation. Similarly, in the case of γ_{19},

	3 Bernoulli terms give	0.67068	88575	83315	$(+9)$
which with the head of the series gives	$\gamma_{19}=0.21630$	66613	69313	(-3)

It is because of this that we used the Tooper multiple precision package, with quadruple precision.

We conclude this paragraph by recording the recurrence relations which we used to calculate the derivatives needed in the Bernoulli terms and in the error estimates.

LEMMA 3. *If* $f_m(x)=(\log x)^m/x$ *then*

$$f^{(r)}(x)=(-1)^r(r!)\,x^{-r-1}\,[(-1)^m\alpha_{0,r}-(-1)^{m-1}\alpha_{1,r}\log x+\ldots+\alpha_{m,r}(\log x)^m]. \tag{7.10}$$

where

$$\begin{cases}\alpha_{0,r+1}=\alpha_{0,r}+\dfrac{1}{r+1}\alpha_{1,r}\\ \alpha_{1,r+1}=\alpha_{1,r}+\dfrac{2}{r+1}\alpha_{2,r}\\ \qquad\ldots\\ \alpha_{m-1,r+1}=\alpha_{m-1,r}+\dfrac{m}{r+1}\alpha_{m,r}\\ \alpha_{m,r}\equiv 1\end{cases} \tag{7.11}$$

with appropriate initial conditions:

$$\alpha_{0,0}=\alpha_{1,0}=\ldots=\alpha_{m-1,0}=0,\ \alpha_{m,0}=1.$$

PROOF. Induction.

Using the same ideas as in our discussion of $f_2^{(r)}$ we can estimate

$$\alpha_{s,r}\doteq\binom{m}{s}(\log r)^{m-s}$$

which gives

$$f_m^{(r)}(x)\doteq(-1)^r r!\,x^{-r-1}[\log x-\log r]^m$$

and indicates that $f_m^{(r)}(x)$ will be of one sign for $x\geq x_0(r)$ where $x_0(r)=O(r)$.

8. Gram's Calculations

Gram [7] gives few details of his computations but we can indicate the lines along which they went. He realized that it was not attractive to extend the calculations of Jensen [9] which were based on the definition

$$\zeta(s)=\lim\left\{\sum_{r=1}^{n}r^{-s}-\frac{n^{1-s}}{1-s}\right\}$$

and sought another approach.

The function

$$\xi(s)=\frac{1}{2}s(s-1)\,\pi^{-s/2}\,\Gamma\left(\frac{1}{2}s\right)\zeta(s) \tag{8.1}$$

is entire and satisfies the functional equation

$$\xi(s)=\xi(1-s).$$

If we write

$$\equiv(t)=\xi(s),\qquad s=\frac{1}{2}+it \tag{8.2}$$

then $\equiv)(t)$ is an even entire function which does not vanish at $t=0$ since $\equiv(0)=\xi\left(\frac{1}{2}\right)=-\frac{1}{8}\pi^{-1/4}\,\Gamma\left(\frac{1}{4}\right)\zeta\left(\frac{1}{2}\right)$. We can write

$$\log\equiv(it)=a_0+a_1t^2+a_2t^4+\ldots$$

where $a_0=\log\xi\left(\frac{1}{2}\right)$. If we can compute the a's, we can, by exponentiating, get the coefficients in the Maclaurin expansion of $\equiv(t)$ and from this, using (8.1) and (8.2) obtain the coefficients in the Laurent expansion of $\zeta(s)$ about $s=1$.

Gram had available Stieltjes' $32D$ values of $\zeta(s)$, $s=2(1)70$, i.e., in principle, the values of $\log\equiv(it)$ for $\pm it=\frac{3}{2}(1)\frac{139}{2}$. The a's are essentially the values of the (even) derivatives of

479-206 O - 73 - 6

$\log \equiv (it)$ at 0 and can be calculated by using appropriate Lagrangian expressions. For instance, we have (see, e.g., [23])

$$f_{1/2}=\frac{1}{256}(3f_{-2}-25f_{-1}+150f_0+150f_1-25f_2+3f_3)-\frac{5}{1024}\mu\delta^6 f_{1/2}+\ldots$$

$$h^2 f''_{1/2}=\frac{1}{48}(-5f_{-2}+39f_{-1}-34f_0-34f_1+39f_2-5f_3)+\frac{259}{5760}\mu\delta^6 f_{1/2}-\ldots..$$

From these expressions we find, when $f_{-n}=f_{n+1}=\log\equiv(i(1/2+n))=\log\xi(n+1)$,

$$a_0\doteq-0.69892\,210,$$

$$a_1\doteq+0.02310\,424.$$

These are to be compared with the values given by Gram

$$a_0=-0.69892\,22679\,45331\,4,$$

$$a_1=+0.02310\,49931\,15419\,0.$$

Gram, actually, made use of independent calculations by himself and Stieltjes of $\zeta(\frac{1}{2})$ which gave a_0 directly and his calculations were more complicated and elaborate than ours, e.g., he used the values of $\log\equiv(t)$ for $it=\frac{1}{2}(1)\frac{29}{2}$.

9. Tooper's Multiple Precision Package

When we realized that we needed multiple precision calculations we were fortunate to have available a package produced by Professor R. F. Tooper. In this we used one storage place for the exponent of our numbers and four for the mantissa (actually it is possible to use up to 19). Besides the usual arithmetic operations included in this package we used the logarithm subroutine. The logarithms were calculated from the usual series

$$\log\frac{1-x}{1+x}=-2\,[x+\tfrac{1}{3}x^3+\tfrac{1}{3}x^5\ldots]$$

after a reduction in the range to $(\frac{1}{2},1)$; a "long division" process was used.

Our preliminary calculations were made on the IBM 360/75 at the California Institute of Technology and the final ones on the IBM 370/155 at the University of South Florida.

We used (7.3), with $n=400$, to compute the γ's and A's, and (3.6) to derive the τ's.

10. Tables

The Stieltjes constants A_r

	(1) Jensen		(2) Gram			
0	0.57721	5665	0.57721	56649	01532	9
1	.07281	5845	.07281	58454	83676	7
2	−.00484	5182	−.00484	51815	96436	1
3	−.00034	2306	−.00034	23057	36717	2
4	.00009	6889	.00009	68904	19394	4

The Stieltjes constants $(-1)^r A_r$—Continued

	(1) Jensen		(2) Gram			
5	−.	6611	−.00006	66110	31810	8
6	−.	0332	−.	03316	24090	9
7	.	0105	.	01046	20945	9
8	−.	0009	−.	00087	33218	1
9			.		94782	7
10			.		56584	2
11			−.		06768	7
12			.		00319	2
13			.		00004	4
14			−.		00002	4
15			.		00000	2

Jensen asserted that only the ninth decimal in his values was doubtful. Gram stated that he believed that his results were correct to 15*D* but could not guarantee this. Our results confirm these statements—the only discrepancies are Jensen's value for A_4 and Gram's value for A_2, A_9.

NEW COMPUTATIONS

r	The gammas γ_r	The Stieltjes constants A_r	The taus τ_{r+1}
0	0.57721 56649 01533	0.57721 56649 01533	0.15986 89037 42431
1	-0.72815 84548 36767 (-1)	0.72815 84548 36767 (-1)	0.65372 59255 88987 (-1)
2	-0.96903 63192 87232 (-2)	-0.48451 81596 43616 (-2)	0.94139 50232 49318 (-2)
3	0.20538 34420 30335 (-2)	-0.34230 57367 17224 (-3)	-0.17996 93810 68913 (-1)
4	0.23253 70065 46730 (-2)	0.96890 41939 44708 (-4)	-0.24514 90765 64097 (-1)
5	0.79332 38173 01063 (-3)	-0.66110 31810 84219 (-5)	-0.16685 79600 04258 (-1)
6	-0.23876 93454 30200 (-3)	-0.33162 40908 75277 (-6)	-0.83440 13692 15834 (-3)
7	-0.52728 95670 57751 (-3)	0.10462 09458 44792 (-6)	0.16416 07577 68938 (-1)
8	-0.35212 33538 03039 (-3)	-0.87332 18100 27380 (-8)	0.28113 16491 35773 (-1)
9	-0.34394 77441 80881 (-4)	0.94782 77782 76236 (-10)	0.27628 25192 88580 (-1)
10	0.20533 28149 09065 (-3)	0.56584 21927 60871 (-10)	0.10215 12232 23534 (-1)
11	0.27018 44395 43904 (-3)	-0.67686 89863 51370 (-11)	-0.24395 66978 16397 (-1)
12	0.16727 29121 05140 (-3)	0.34921 15936 67203 (-12)	-0.68551 96259 04596 (-1)
13	-0.27463 80660 37602 (-4)	0.44104 24741 75775 (-14)	-0.10325 15678 59017
14	-0.20920 92620 59300 (-3)	-0.23997 86221 77100 (-14)	-0.96775 02330 08802 (-1)
15	-0.28346 86553 20241 (-3)	0.21677 31220 07268 (-15)	-0.90996 45657 46961 (-2)
16	-0.19969 68583 08970 (-3)	-0.95444 66076 36696 (-17)	0.19238 77261 64982
17	0.26277 03710 99183 (-4)	-0.73876 76660 53864 (-19)	0.49962 39621 32792
18	0.30736 84081 49253 (-3)	0.48008 50782 48807 (-19)	0.80971 09289 54417
19	0.50360 54530 47356 (-3)	-0.41399 56737 71331 (-20)	0.85772 10307 17131

We are indebted to Peter Weinberger and Herman P. Robinson for detailed comments on our tables and especially to Henry C. Thacher, Jr. for showing us his manuscript [26] which includes a table of γ_r, $r=0(1)35$, 44 to 28 S as well as the coefficients of $z\zeta(1+z)$ together with Chebyshev series for this function in the ranges $[-1/2, 1/2]$ and $[0, 1]$. His values were also obtained by the Euler-Maclaurin method and our values agree with his up to the last digit except as indicated: specifically Thacher has for γ_8 . . . 9|509 . . . and for γ_9 . . . 0|481 . . . where we have . . . 9| and . . . 1|.

11. References

[1] Briggs, W. E., Some constants associated with the Riemann zeta-function, Michigan Math. J. **31**, 117–121 (1955–6).
[2] Briggs, W. E., and Chowla, S., The power series coefficients of $\zeta(s)$, American Math. Monthly **62**, 323–325 (1955).
[3] Correspondance d'Hermite et de Stieltjes, I, II (Gauthier-Villars, Paris, 1905).
[4] Dahlquist, G., Gustafson, S. A., and Siklósi, K., Convergence acceleration from the point of view of linear programming, BIT, **5**, 1–16 (1965).
[5] Francis, E. C., and Littlewood, J. E., Examples in infinite series with solutions (Cambridge, 1928).
[6] Glaisher, J. W. L., On Dr. Vacca's series for γ. Quarterly Journal Math., **41**, 365–368 (1910).
[7] Gram, J. P., Note sur le calcul de la fonction $\zeta(s)$ de Riemann. Oversigt. K. Danske Vidensk. (Selskab Forhandlingar, 305–308 1895).
[8] Hardy, G. H., Note on Dr. Vacca's series for γ. Quarterly Journal Math., **43**, 215–216 (1912). (= Collected Papers, IV (1969), 475–476, Oxford).
[9] Jensen, J. L. W. V., Sur la fonction $\zeta(s)$ de Riemann. Comptes Rendus (Paris), **104**, 1156–1159 (1887).
[10] Knopp, K., Theory and application of infinite series, tr. R. C. Young (Blackie, London and Glasgow, 1928).
[11] Knuth, D. E., Euler's Constant to 1271 places. Math. Comp., **16**, 275–281 (1962).
[12] Ramanujan, S., Collected papers (Cambridge, 1927).
[13] Rosser, J. Barkley, Transformations to speed the convergence of series, J. Res. Nat. Bur. Stand., **46**, 56–64 (1951) RP 2175.
[14] Todd, John, Optimal parameters in two programs. B. Inst. Math. **6**, 31–35 (1970).
[15] Vacca, G., A new series for the Eulerian constant .577 . . .. Quarterly Journal Math., **41**, 363–364 (1910).
[16] Wheelon, A. D., Table of Summable Series and Integrals Involving Bessel Functions, p. 41 (Holden Day, San Francisco, 1968).
[17] Wilkinson, J. H., The algebraic eigenvalue problem, p. 247 (Oxford, 1965).
[18] Ostrowski, A. M., Note on Poisson's treatment of the Euler-Maclaurin formula, Comment. Math. Helv. **44**, 202–206 (1969).
[19] Ostrowski, A. M., On the remainder terms of the Euler-Maclaurin formula, J. Reine Angew. Math. **239/240**, 268–286 (1970).
[20] Boas, R. P. Jr., and Wrench, J. W. Jr., Partial sums of the harmonic series, Amer. Math. Monthly **78**, 864–870 (1971).
[21] Lammel, E., Über die Koeffizienten in der Reihenentwicklung von $(s-1)\}(s)$ nach Potenzen von $s-1$, Nachrichten d. Osterreichischen Math. Gesell. **91**, 41 (1970).
[22] Hardy, G. H., Pure Mathematics (Cambridge, 1925).
[23] H. M. Nautical Almanac Office, Interpolation and Allied Tables (London, 1956).
[24] Jacobstahl, E., (Ed. S. Selberg). Über die Eulersche Konstante (Norske Vid. Selsk. Skr., Trondheim, 1967, No. 11, 24 pp., 1968).
[25] Addison, A. W., A series representation for Euler's constant. Amer. Math. Monthly **74**, 823–824 (1967).
[26] Henry C. Thacher, Jr., On expansions of the Riemann Zeta function, to appear.

(Paper 76B3&4–369)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Vol. 76B, Nos. 3 and 4, July–December 1972

Confidence Limits for the Abscissa of Intersection of Two Linear Regressions

James J. Filliben

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

and

John E. McKinney

Institute for Materials Research, National Bureau of Standards, Washington, D.C. 20234

(August 25, 1972)

A method to determine the confidence limits for the abscissa of the intersection of two linear regressions has been developed. This method does not require the assumption of equal variance for the two regressions, as was necessary with previous methods. A numerical example is included on thermodynamic, glass transition data for which this method is applicable. Comparisons are made between the results using equal and unequal variance assumptions. A FORTRAN subroutine is included for computations using both assumptions.

Key words: Abscissa; confidence limits; glass transition; intercept; intersection; linear; regression; second order transition; statistics; variance.

1. Introduction

In this paper confidence limits are derived for the abscissa of the intersection point of two linear regressions. As illustrated in figure 1, it is assumed that there are independent realizations of two linear regressions at ordered values of a single independent variable X. The problem is to determine confidence limits for the abscissa X_0 of the intersection point of the two linear regressions.

This problem was originally considered and solved by Fisher [1][1] and later by Kastenbaum [2], but under the assumption of a common error variance for the two regression lines. Hinkley [3] derived the asymptotic distribution of the maximum likelihood estimator for the intersection point abscissa of two linear regressions, also assuming equal variances. Robison [4] presented estimators and confidence limits for the intersection point abscissa of two polynomial regressions under the equal variance assumption. Hudson [5] derived least squares estimators (but no confidence limits) for the intersection point abscissas of two or more general regression models for the unequal variance case.

The principal contribution of the present paper is a derivation of confidence limits for the intersection point abscissa without resorting to the equal variance assumption. Although only the linear regression case is considered below, extension of results to more complicated regression models is possible.

The results of this derivation have direct application to thermodynamic data to evaluate the uncertainties in the second order transition temperature T_2 and the glass transition temperature T_g.

AMS Subject Classification: 62.55.

[1] Figures in brackets indicate the literature references at the end of this paper.

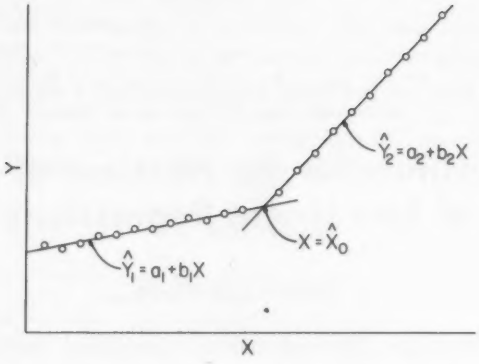


FIGURE 1. *Intersection of two linear regressions.*

The second order transition is an equilibrium phenomenon, whereas the glass transition is a non-equilibrium one for which T_g depends upon the thermodynamic history by which the glass is formed. Both transitions may be manifested by discontinuities in the second order properties (for example, thermal expansivity or heat capacity) with respect to temperature. The corresponding first order properties (volume or enthalpy) vary with temperature in a manner similar to Y versus X in figure 1 for which the transition temperature may be determined by the intersection of the two straight lines representative of the response in the two regions. Accordingly, this method may be applied directly to any of the first order properties to determine the confidence bands of the transition temperature as defined[2] from the intersection of the two regression lines. The method is sufficiently general that it is applicable to cases where data are not available in the proximity of the intersection, which, of course, leads to a less accurate prediction of its abscissa.

Two methods are described to evaluate the confidence limits for the abscissa of intersection. These are based on the assumptions of equal and unequal variances for the regression lines determined from the two sets. The advantage of the equal variance case is that the underlying assumption of equal variance permits a relatively simple development of the method, whereas the advantage of the unequal variance case is, of course, in its more general application. The ANSI FORTRAN subroutine to calculate the confidence limits for both cases is given in the appendix, and numerical examples are included in later discussions.

2. Derivation of Confidence Limits

As illustrated in figure 1, let $\hat{Y}_1(X) = a_1 + b_1X$ and $\hat{Y}_2(X) = a_2 + b_2X$ be two fitted least squares regressions based on n_1 and n_2 observations $(X_{ij}, i=1, 2; j=1, 2, \ldots, n_i)$ respectively, where the random variables $\hat{Y}_i(X)$ are assumed to be normally and independently distributed about the true unknown regression lines $\mu_i(X) = \alpha_i + \beta_iX$ with variances σ_i^2. (α_i, β_i, and σ_i^2 are true, but unknown, parameter values.)

The problem is to determine confidence limits for the abscissa X_0 of the intersection point (X_0, Y_0) of the two unknown regression lines.

The starting point for our derivation is the fact that a_i and b_i, the usual least squares estimators, are linear combinations of normally distributed variables and hence themselves are normally distributed. Thus for all X, $\hat{Y}_i(X) = a_i + b_iX$ is also normally distributed, as is the difference $\hat{Y}_1(X) - \hat{Y}_2(X)$. Since $\hat{Y}_1(X) - \hat{Y}_2(X)$ has mean $\mu_1(X) - \mu_2(X)$ and variance

[2] This definition of the glass temperature is one of several which are not necessarily equivalent.

$\sigma_1^2 V_1(X) + \sigma_2^2 V_2(X)$, where

$$V_i(X) = \frac{1}{n_i} + \frac{(X - \bar{X}_i)^2}{S_i^2}, \qquad i = 1, 2$$

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, \qquad i = 1, 2 \tag{1}$$

$$S_i^2 = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2, \qquad i = 1, 2,$$

then it follows that all X

$$\frac{[\hat{Y}_1(X) - \hat{Y}_2(X)] - [\mu_1(X) - \mu_2(X)]}{\sqrt{\sigma_1^2 V_1(X) + \sigma_2^2 V_2(X)}} \tag{2}$$

has a standardized (mean 0, variance 1) normal distribution.

Consider first of all the equal variance case ($\sigma_1^2 = \sigma_2^2 = \sigma^2$, say), treated earlier [1]. It is seen that eq (2) simplifies to

$$\frac{[\hat{Y}_1(X) - \hat{Y}_2(X)] - [\mu_1(X) - \mu_2(X)]}{\sigma \sqrt{V_1(X) + V_2(X)}}. \tag{3}$$

Thus

$$\frac{[\hat{Y}_1(X) - \hat{Y}_2(X)] - [\mu_1(X) - \mu_2(X)]}{s \sqrt{V_1(X) + V_2(X)}} \tag{4}$$

has a t distribution with $n_1 + n_2 - 4$ degrees of freedom, where s is the square root of the pooled unbiased estimator of σ^2:

$$s = \sqrt{\frac{(n_1 - 2)s_1^2 + (n_2 - 2)s_2^2}{n_1 + n_2 - 4}}, \tag{5}$$

and where s_i^2 is the unbiased estimator for σ_i^2 given by the residual variance obtained from the least squares fit of line i.

The quantity (4) is distributed as t for all X, and in particular for $X = X_0$ in which case $\mu_1(X) - \mu_2(X)$ vanishes. It then follows that for every probability $p\,(0 < p < 1)$

$$\text{Prob}\left\{\left(\frac{\hat{Y}_1(X_0) - \hat{Y}_2(X_0)}{s\sqrt{V_1(X_0) + V_2(X_0)}}\right)^2 \leq \left(G_t\left(\frac{1+p}{2}\right)\right)^2\right\} = p \tag{6}$$

where $G_t\left(\frac{1+p}{2}\right)$ is the $100\left(\frac{1+p}{2}\right)$ percent point of the t distribution with $n_1 + n_2 - 4$ degrees of freedom. The squared numerator $[\hat{Y}_1(X_0) - \hat{Y}_2(X_0)]^2$ and squared denominator $s^2[V_1(X_0) + V_2(X_0)]$ above are both quadratics in X_0. Thus eq (6) is of the form

Prob $\left\{\frac{Q_1}{Q_2} \leq G^2\right\}$ (or equivalently Prob $\{Q_1 - G^2 Q_2 \leq 0\}$) where Q_i denote quadratic expressions, and G denotes $G_t\left(\frac{1+p}{2}\right)$. Since $Q_1 - G^2 Q_2$ is itself a quadratic (say $AX_0^2 + BX_0 + C$), then the

desired confidence interval consists of those X_0 such that $AX_0^2 + BX_0 + C \leq 0$ and the desired confidence limits for X_0 are simply the two roots of $AX_0^2 + BX_0 + C = 0$.

The coefficients A, B, and C are rather tediously derived; omitting the intermediate algebra, these coefficients are

$$A = (b_1 - b_2)^2 - s^2G^2\left[\frac{1}{S_1^2} + \frac{1}{S_2^2}\right]$$

$$B = 2(a_1 - a_2)(b_1 - b_2) + 2s^2G^2\left[\frac{\bar{X}_1}{S_1^2} + \frac{\bar{X}_2}{S_2^2}\right] \tag{7}$$

$$C = (a_1 - a_2)^2 - s^2G^2\left[\frac{1}{n_1} + \frac{1}{n_2} + \frac{\bar{X}_1^2}{S_1^2} + \frac{\bar{X}_2^2}{S_2^2}\right].$$

It is of interest to note that the midpoint $-B/2A$ of the confidence interval is not identical to the maximum likelihood estimator

$$\hat{X}_0 = -\frac{a_1 - a_2}{b_1 - b_2}, \tag{8}$$

however, the two estimators converge to a common limit as $s^2 \to 0$. For a less detailed derivation of the above confidence interval results, see references [1–3].

For the more general $\sigma_1^2 \neq \sigma_2^2$ case, the distributional complications which arise are identical to those encountered in the Behrens-Fisher 2-means problem [6]. The solution outlined below is an extension of the Welch-Aspin [7, 8, 9] solution to the Behrens-Fisher problem. Proceeding in a manner analogous to Welch and Aspin, we see that for all X,

$$\frac{[\hat{Y}_1(X) - \hat{Y}_2(X)] - [\mu_1(X) - \mu_2(X)]}{\sqrt{s_1^2V_1(X) + s_2^2V_2(X)}} \tag{9}$$

has approximately a t distribution with ν degrees of freedom, where ν is given by

$$\frac{1}{\nu} = \frac{1}{n_1 - 2}\left(\frac{V_1(X_0)}{V_1(X_0) + V_2(X_0)}\right)^2 + \frac{1}{n_2 - 2}\left(\frac{V_2(X_0)}{V_1(X_0) + V_2(X_0)}\right)^2 \tag{10}$$

A minor complication at this point is the fact this X_0 in (10) is an unknown parameter. In order to arrive at an approximate value for ν the maximum likelihood estimate $\hat{X}_0$ (as defined in (8)) is used. The development now parallels the $\sigma_1^2 = \sigma_2^2$ case, and so again 100 p confidence limits for X_0 are obtained by solving a quadratic equation, but with the following slightly-modified coefficients:

$$A = (b_1 - b_2)^2 - G^2\left[\frac{s_1^2}{S_1^2} + \frac{s_2^2}{S_2^2}\right]$$

$$B = 2(a_1 - a_2)(b_1 - b_2) + 2G^2\left[\frac{s_1^2\bar{X}_1}{S_1^2} + \frac{s_2^2\bar{X}_2}{S_2^2}\right] \tag{11}$$

$$C = (a_1 - a_2)^2 - G^2\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} + \frac{s_1^2\bar{X}_1^2}{S_1^2} + \frac{s_2^2\bar{X}_2^2}{S_2^2}\right]$$

Note that although the unequal variance case basically parallels the equal variance case, the former does not numerically reduce to the latter when $s_1^2 = s_2^2$. The confidence intervals usually are distinct even when $s_1^2 = s_2^2$. In deciding whether to use the equal variance results or the unequal variance results, there are two considerations to be taken into account: physical theory and the variance ratio F test. The general rule of thumb is to always use the unequal variance results unless the following two conditions are simultaneously satisfied: (1) there is a firm physical reason for assuming equal variances; and (2) the F ratio S_1^2/S_2^2 is not significantly different from unity. This recommended procedure is at times conservative, but more important, it is less assumption-dependent and hence more generally valid.

In the appendix, an ANSI FORTRAN subroutine is presented which first of all tests to see whether the equal variance assumption is tenable, and secondly computes confidence limits for X_0 for both the equal variance and the unequal variance cases. The subroutine is double precision in input, internal operation, and output. Definitions of input parameters and instructions for the use of the subroutine are given in the comment statements within the subroutine.

3. Example: Calculation of the 95 percent Confidence Limits for the Pressure-Dependent Glass Temperature

As an example, the 95 percent confidence limits have been calculated for the pressure-dependent glass transition temperature of poly(vinyl acetate), an amorphous, glass-forming polymer. The PVT (pressure-volume-temperature) data for the glass, designated as set 1, are given in table 1, and the data for the liquid, designated as set 2, are given in table 2. Both sets of data,

TABLE 1. *Specific volume values for the glassy region*

T/P	0	100	200	300	400	500	600	700	800
−30	0.82884	0.82654	0.82438	0.82225	0.82017	0.81816	0.81613	0.81419	0.81237
−25	.83002	.82768	.82547	.82334	.82121	.81914	.81715	.81516	.81327
−20	.83110	.82874	.82648	.82430	.82218	.82010	.81808	.81606	.81415
−15	.83238	.82997	.82767	.82549	.82331	.82117	.81919	.81717	.81522
−10	.83350	.83117	.82886	.82663	.82443	.82224	.82022	.81817	.81620
−5	.83467	.83233	.83003	.82774	.82557	.82340	.82131	.81933	.81729
0	.83581	.83351	.83115	.82889	.82666	.82447	.82237	.82033	.81825
5	.83706	.83478	.83241	.83012	.82794	.82569	.82355	.82140	.81941
10	.83812	.83593	.83359	.83130	.82903	.82679	.82465	.82248	.82040
15	.83937								
20	.84061								

Units: T in °C, P in Bars, v in cm³/g.

TABLE 2. *Specific volume data for the liquid region*

T/P	0	100	200	300	400	500	600	700	800
35	0.84572	0.84148							
40	.84870	.84444	0.84041	0.83670					
45	.85174	.84733	.84321	.83921	0.83549	0.83184			
50	.85486	.85041	.84608	.84206	.83817	.83447	0.83087	0.82757	
55	.85791	.85349	.84894	.84491	.84092	.83709	.83340	.82988	0.82666
60	.86104	.85628	.85179	.84769	.84367	.83980	.83607	.83256	.82904
65	.86407	.85933	.85472	.85052	.84648	.84250	.83874	.83523	.83174
70	.86723	.86218	.85755	.85324	.84913	.84507	.84108	.83734	.83378
75	.87038	.86536	.86038	.85594	.85171	.84762	.84367	.84000	.83626
80	.87343	.86829	.86342	.85881	.85453	.85036	.84641	.84257	.83889
85	.87669	.87140	.86636	.86169	.85728	.85308	.84911	.84532	.84155
90	.87986	.87438	.86923	.86444	.86000	.85574	.85172	.84776	.84396
95	.88301	.87741	.87207	.86722	.86269	.85823	.85424	.85025	.84641
100	.88622	.88043	.87500	.87012	.86534	.86095	.85678	.85284	.84904

Units: T in °C, P in bars, v in cm³/g

taken from reference [10], are illustrated in figure 2, where the specific volume v taken from the above tables is plotted as a function of temperature at the different pressures indicated. The specific volume data $v_1(T, P)$ for the glass (at temperatures between −30 and 20 °C) were obtained after the glass was formed at atmospheric pressure at a constant cooling rate of 5 °C/h commencing at equilibrium in the liquid region. The data $v_2(T, P)$ for the liquid (35 to 100 °C) are taken to be in true equilibrium. A characteristic of a glass is its large increase in its viscoelastic relaxation times with decreasing temperature. The glass transition is a manifestation of the increase of relaxation times during formation of the glass. Although the data shown here for the glass are not representative of the attainment of true equilibrium, the relaxation times at these temperatures are very long in comparison to experimental times. Accordingly, the glass data may be treated as being in apparent equilibrium (pseudoequilibrium) over any practical time scale, as has been done in this case.

Since the volume isobars vary nearly linearly with temperature in both the liquid and glass regions, and since the slopes for each isobar in these regions are distinct, the results of section 2 are applicable. By treating each of the two regions separately, each of the volume isobars may be represented by two distinct regression lines of the form

$$\hat{v}_i = a_i + b_i T, \qquad i = 1, 2. \tag{12}$$

The intersection of these two regression lines at each pressure gives the "maximum likelihood estimator"

$$\hat{T}_g = -\frac{a_1 - a_2}{b_1 - b_2}, \tag{13}$$

which is taken here as an estimate of the true, but unknown, glass temperature T_g for the particular thermodynamic history by which the glass was formed. In figure 2 the regression lines of the form of eq (12) are shown by the solid lines with positive slope at each pressure. The solid line

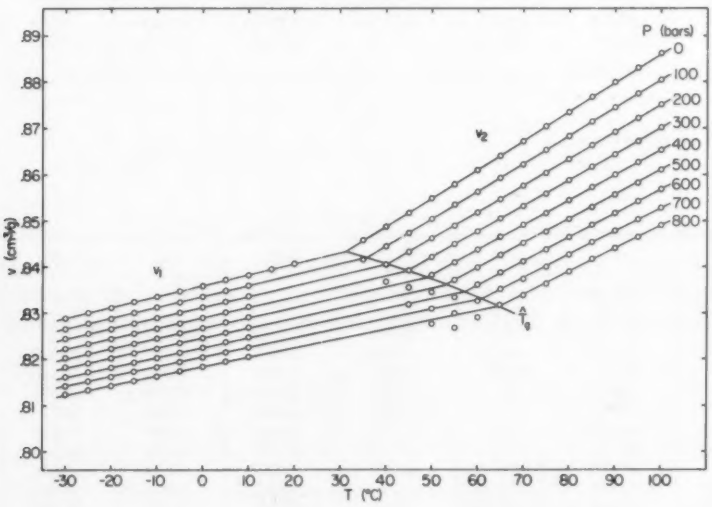


FIGURE 2. *Specific volume of poly(vinyl acetate) as a function of temperature at different pressures.*

An estimate of the glass transition is obtained from the intersection of the volume isobars taken to be linear with temperature.

with negative slope gives the positions of $\hat{T}_g$, which depend strongly on the pressure. It is seen that some of the data representative of the liquid region fall to the left of $\hat{T}_g$. Although this result may be paradoxical, it is simply a consequence of the fact that the liquid data are equilibrium data, and the glass are pseudoequilibrium data[4] resulting from the particular thermodynamic history by which the glass was formed.

Ninety-five percent confidence limits for T_g were calculated using both the equal and unequal variance assumptions. The paired statistical quantities necessary to calculate the confidence limits are given in tables 3 and 4 for the glass and liquid, respectively, where P is the pressure, a_i is the intercept, b_i is the slope, s_i is the standard deviation of the residuals about the regression lines,

TABLE 3. *Statistical data for the Glassy Region*

P	a_1	$b_1 \times 10^3$	$s_1 \times 10^4$	n_1	$\bar{T}_1$	S_1^2
0	0.835843	0.23469	0.495	11	−5.0	2750.0
100	.833531	.23590	.565	9	−10.0	1500.0
200	.831192	.23123	.778	9	−10.0	1500.0
300	.828928	.22660	.855	9	−10.0	1500.0
400	.826716	.22270	.910	9	−10.0	1500.0
500	.824512	.21713	1.021	9	−10.0	1500.0
600	.822417	.21327	.793	9	−10.0	1500.0
700	.820328	.20860	.737	9	−10.0	1500.0
800	.818301	.20270	.897	9	−10.0	1500.0

TABLE 4. *Statistical data for the Liquid Region*

P	a_2	$b_2 \times 10^3$	$s_2 \times 10^4$	n_2	$\bar{T}_2$	S_2^2
0	0.823686	0.62355	1.146	14	67.5	5687.5
100	.820417	.59934	.869	14	67.5	5687.5
200	.817195	.57760	.683	13	70.0	4550.0
300	.814189	.55834	.832	13	70.0	4550.0
400	.811039	.54362	.565	12	72.5	3575.0
500	.808010	.52946	.589	12	72.5	3575.0
600	.804836	.51987	.952	11	75.0	2750.0
700	.802096	.50698	1.341	11	75.0	2750.0
800	.799213	.49703	1.564	10	77.5	2062.5

n_i is the number of observations, $\bar{T}_i$ is the average temperature, and S_i^2 is the sum of the squared temperature deviations:

$$S_i^2 = \sum_{j=1}^{n_i} (T_{ij} - \bar{T}_i)^2, \qquad i = 1, 2.$$

The results of these calculations are summarized in table 5 for both the equal and unequal variance cases. G is the 97.5 percent point of the t_ν distribution with ν degrees of freedom where $\nu = n_1 + n_2 - 4$ in the equal variance case, and where ν is defined by eq (10) in the unequal variance case. $-A/2B$ is the midpoint of the confidence interval where A and B are defined by eqs (7) in the equal variance case and eqs (11) in the unequal variance case. ΔT is the 95 percent confidence interval width which is equal to the difference of the two roots of the quadratic equation for which the constants are defined in eqs (7) and (11) for equal variance and unequal variance, respectively. s_1/s_2 is the ratio of residual standard deviations, and R is the ratio of ΔT for the equal variance case to ΔT for the unequal variance case. Perusal of table 5 reveals several interesting points. First, it is noted that the corresponding midpoints $-B/2A$ are essentially identical for the equal

[4] In principle, at least at temperatures over a limited range below T_g shown here, the specific volume values for the glass will relax to extrapolations of the liquid isobars shown provided sufficient time is allowed. Since this time may be as long as several million years (depending on the temperature and pressure), the attainment of true equilibrium in all cases (as with the liquid values shown) is not practical.

TABLE 5. *Estimated values of* T_g *and 95 percent Confidence Intervals*

P	$\hat{T}_g$	Equal variance				Unequal variance				$\frac{s_1}{s_2}$	R
		ν	G	$-B/2A$	ΔT	ν	G	$-B/2A$	ΔT		
0	31.26	21	2.08	31.27	0.92	17	2.11	31.26	0.79	0.432	1.162
100	36.08	19	2.09	36.09	1.18	9	2.26	36.09	1.02	0.650	1.158
200	40.41	18	2.10	40.42	1.26	9	2.26	40.42	1.44	1.138	0.874
300	44.43	18	2.10	44.44	1.62	8	2.31	44.44	1.80	1.028	0.898
400	48.85	17	2.11	48.86	1.56	8	2.31	48.87	2.07	1.609	0.752
500	52.84	17	2.11	52.85	1.84	8	2.31	52.86	2.53	1.734	0.729
600	57.34	16	2.12	57.36	2.24	8	2.31	57.35	2.21	.833	1.014
700	61.10	16	2.12	61.13	3.03	7	2.36	61.11	2.34	.549	1.296
800	64.85	15	2.13	64.89	3.77	7	2.36	64.87	3.02	.573	1.248

and unequal variance cases, and these, in turn, are essentially identical to the maximum likelihood estimator $\hat{T}_g$. As mentioned earlier these quantities converge to the common limit $\hat{T}_g$ as $s_i^2 \to 0$. Secondly, the confidence limits vary considerably between the equal and unequal variance cases (the ratio R being at times both greater than or less than unity). In the case where s_1 and s_2 are nearly equal ($P = 300$ bars), R differs considerably from unity; and in the case where R is nearly unity ($P = 700$ bars), s_1 and s_2 are considerably distinct. This behavior illustrates the fact that the equal variance case is not a special case of the unequal variance case as might be expected. Thirdly, there appears to be little correlation between the ratio of confidence interval widths and the ratio of residual standard deviations.

The values of the maximum likelihood estimator $\hat{T}_g$ and those for the 95 percent confidence interval width ΔT (each centered about its midpoint $-B/2A$) for both the equal variance ($\sigma_1^2 = \sigma_2^2$) and the unequal variance ($\sigma_1^2 \neq \sigma_2^2$) cases are shown on figure 3 as functions of pressure. The median solid line is a quadratic regression of $\hat{T}_g$ with respect to pressure. The two extreme solid lines are quadratic regressions of the confidence limits with respect to pressure for the unequal variance case only. It is clear that a trend exists between the confidence interval and pressure. The principal reason for the divergence of the confidence interval with increasing pressure is that the difference $(\hat{T}_g - \bar{T}_1)$ between the estimated glass temperature $\hat{T}_g$ and the average temperature for the glass $\bar{T}_1$ also increases considerably with pressure as seen from figure 2. Thus is apparent the importance of obtaining data in the vicinity of the glass temperature (or, in general, the intersection) in order to obtain an accurate estimate.

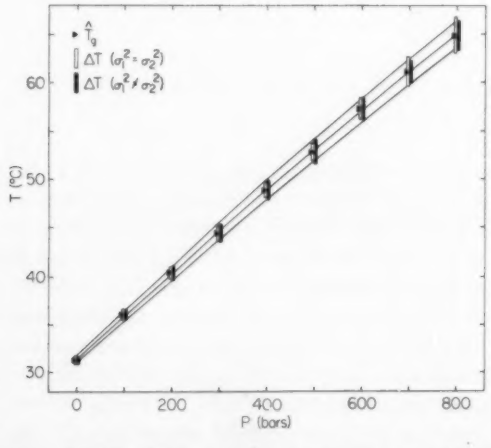


FIGURE 3. *Maximum Likelihood Estimator,* $\hat{T}_g$, *and 95 percent Confidence Intervals as functions of pressure.*

The solid lines are quadratic regressions of these quantities.

```
      SUBROUTINE ABSCIS(AN1,AN2,XBAR1,XBAR2,SSQDX1,SSQDX2,ALP1,ALP2,BET1
     1,BET2,RSD1,RSD2)
C     THE PURPOSE OF THIS ROUTINE IS TO COMPUTE 95 PERCENT CONFIDENCE LIMITS
C     FOR THE ABSCISSA OF THE INTERSECTION POINT OF TWO LINEAR REGRESSIONS
C     FOR BOTH THE EQUAL VARIANCES CASE AND THE UNEQUAL VARIANCES CASE
C     AN1 = THE NUMBER OF (X,Y) PAIRS FOR LINE 1
C     AN2 = THE NUMBER OF (X,Y) PAIRS FOR LINE 2
C     XBAR1 = THE SAMPLE MEAN OF THE X VALUES FOR LINE 1
C     XBAR2 = THE SAMPLE MEAN OF THE X VALUES FOR LINE 2
C     SSQDX1 = THE SUM OF THE SQUARED DEVIATIONS OF THE X VALUES FOR LINE 1
C              ABOUT XBAR1
C     SSQDX2 = THE SUM OF THE SQUARED DEVIATIONS OF THE X VALUES FOR LINE 2
C              ABOUT XBAR2
C     ALP1 = THE Y INTERCEPT OF LINE 1
C     ALP2 = THE Y INTERCEPT OF LINE 2
C     BET1 = THE SLOPE OF LINE 1
C     BET2 = THE SLOPE OF LINE 2
C     RSD1 = THE RESIDUAL STANDARD DEVIATION OF LINE 1
C     RSD2 = THE RESIDUAL STANDARD DEVIATION OF LINE 2
C     THIS ROUTINE IS DOUBLE PRECISION IN INPUT, INTERNAL OPERATION, AND OUTPUT
      DIMENSION Z(30)
      DOUBLE PRECISION AN1,AN2,XBAR1,XBAR2,SSQDX1,SSQDX2
      DOUBLE PRECISION ALP1,ALP2,BET1,BET2,RSD1,RSD2
      DOUBLE PRECISION Z
      DOUBLE PRECISION P,S,DELALP,DELBET,AML
      DOUBLE PRECISION ANUEQ,ANUNEQ,GEQ,GNEQ,PP,HPP
      DOUBLE PRECISION RATIO,AM,AN,HOLD
      DOUBLE PRECISON A1,A2,A3,A,B1,B2,B3,B,C1,C2,C3,C
      DOUBLE PRECISION DISC,ROOT1,ROOT2,AMID,WIDTH1,WIDTH2,RATWID
      DOUBLE PRECISION DSQRT
      DATA Z(1),Z(2),Z(3),Z(4),Z(5),Z(6),Z(7),Z(8),Z(9),Z(10),Z(11),
     1Z(12),Z(13),Z(14),Z(15),Z(16),Z(17),Z(18),Z(19),Z(20),Z(21),
     1Z(22),Z(23),Z(24),Z(25),Z(26),Z(27),Z(28),Z(29),Z(30)/12.7026D0,
     14.3027D0,3.1824D0,2.7764D0,2.5706D0,12.4469D0,2.3646D0,2.3060D0,
     12.2622D0,2.2281D0,2.2010D0,2.1788D0,12.1604D0,2.1448D0,2.1315D0,
     12.1199D0,2.1096D0,2.1009D0,2.0930D0,12.0860D0,2.0796D0,2.0739D0,
     12.0687D0,2.0639D0,2.0595D0,2.0555D0,12.0518D0,2.0484D0,2.0452D0,
     12.0423D0/
C
C     COMPUTE NEEDED CONSTANTS
C
      P=.95D0
      S=DSQRT(((AN1-2.0D0)*RSD1*RSD1+(AN2-2.0D0)*RSD2*RSD2)/(AN1+AN2-4.0
     1D0))
      DELALP=ALP1-ALP2
      DELBET=BET1-BET2
C
C     COMPUTE THE MAXIMUM LIKELIHOOD ESTIMATOR
O
      AML=-DELALP/DELBET
C
C     COMPUTE THE APPROPRIATE DEGREES OF FREEDOM FOR THE T PERCENT POINTS
C     FOR THE EQUAL VARIANCES CASE AND FOR THE UNEQUAL VARIANCES CASE
C
      ANUEQ=AN1+AN2-4.0D0
      V1=(1.0D0/AN1)+(AML-XBAR1)*(AML-XBAR1)/SSQDX1
      V2=(1.0D0/AN2)+(AML-XBAR2)*(AML-XBAR2)/SSQDX2
      ANUNEQ=(1.0D0/(AN1-2.0D0))*((V1/(V1+V2))**2)+(1.0D0/(AN2-2.0D0))*(
     1(V2/(V1+V2))**2)
      ANUNEQ=1.0D0/ANUNEQ
```

```
      NUEQ=ANUEQ+0.1D0
      NUNEQ=ANUNEQ+0.1D0
      IF(NUEQ.LE.30)GEQ=Z(NUEQ)
      IF(NUEQ.GT.30)GEQ=1.96D0*DSQRT(ANUEQ/(ANUEQ-2.0D0))
      IF(NUNEQ.LE.30)GNEQ=Z(NUNEQ)
      IF(NUNEQ.GT.30)GNEQ=1.96D0*DSQRT(ANUNEQ/(ANUNEQ-2.0D0))
      PP=(P+1.0D0)/2.0D0
      HPP=100.0D0*PP
C
C     WRITE OUT THE INPUT DATA
C     IT IS ASSUMED THAT THE OUTPUT UNIT IS 6.  IF THIS IS NOT THE CASE, THEN
C     CHANGE THE NEXT STATEMENT APPROPRIATELY.
C
      IP=6
      WRITE(IP,998)
      WRITE(IP,170)
      WRITE(IP,999)
      WRITE(IP,180)
      WRITE(IP,999)
      WRITE(IP,190)AN1,AN2
      WRITE(IP,200)XBAR1,XBAR2
      WRITE(IP,210)SSQDX1,SSQDX2
      WRITE(IP,220)ALP1,ALP2
      WRITE(IP,230)BET1,BET2
      WRITE(IP,240)RSD1,RSD2
      WRITE(IP,999)
      WRITE(IP,250)P
      WRITE(IP,999)
      WRITE(IP,260)HPP,NUEQ,GEQ
      WRITE(IP,270)HPP,NUNEQ,GNEQ
C
C     COMPUTE THE RATIO OF THE LARGER ESTIMATED RESIDUAL VARIANCE TO
C     THE SMALLER ESTIMATED RESIDUAL VARIANCE
C
      RATIO=(RSD1*RSD1)/(RSD2*RSD2)
      AM=AN1-2.0D0
      AN=AN2-2.0D0
      IF(RATIO.GT.1.0D0)GOTO110
      RATIO=1.0D0/RATIO
      HOLD=AM
      AM=AN
      AN=HOLD
110   WRITE(IP,999)
      M1=AM+0.1
      N1=AN+0.1
      WRITE(IP,280)RATIO
      WRITE(IP,310)M1,N1
      WRITE(IP,999)
      WRITE(IP,290)AML
      WRITE(IP,999)
      WRITE(IP,300)S
      WRITE(IP,999)
      WRITE(IP,999)
C
C     COMPUTE THE COEFFICIENTS OF THE QUADRATIC EQUATION FOR THE EQUAL
C     VARIANCES CASE
C
      WRITE(IP,340)
      WRITE(IP,999)
      A1=DELBET**2
      A2=S*S*GEQ*GEQ
```

```
      A3=(1.0D0/SSQDX1)+(1.0D0/SSQDX2)
      B1=2.0D0*DELALP*DELBET
      B2=2.0D0*A2
      B3=(XBAR1/SSQDX1)+(XBAR2/SSQDX2)
      C1=DELALP**2
      C2=A2
      C3=(1.0D0/AN1)+(1.0D0/AN2)+(XBAR1*XBAR1/SSQDX1)+(XBAR2*XBAR2/SSQDX
     12)
      A=A1-A2*A3
      B=B1+B2*B3
      C=C1-C2*C3
C
C     CHECK FOR ZERO COEFFICIENT OF THE QUADRATIC TERM FOR THE EQUAL
C     VARIANCES CASE
C
      IF(A.NE.0.0D0)GOTO120
      AMID=-C/B
      WRITE(IP,320)
      WRITE(IP,321)AMID
      GOTO140
C
C     COMPUTE THE ROOTS OF THE QUADRATIC, AND THE MIDPOINT OF THE ROOTS
C     FOR THE EQUAL VARIANCES CASE
C
  120 DISC=B*B-4.0D0*A*C
C
C     CHECK FOR A NEGATIVE DISCRIMINANT FOR THE EQUAL VARIANCES CASE
C
      IF(DISC.GE.0.0D0)GOTO130
      WRITE(IP,330)DISC
      GOTO140
  130 ROOT1=(-B+DSQRT(DISC))/(2.0D0*A)
      ROOT2=(-B-DSQRT(DISC))/(2.0D0*A)
      AMID=-B/(2.0D0*A)
      WIDTH1=ROOT1-ROOT2
      WRITE(IP,350)ROOT2
      WRITE(IP,360)ROOT1
      WRITE(IP,370)AMID
      WRITE(IP,380)WIDTH1
      WRITE(IP,999)
      WRITE(IP,999)
C
C     COMPUTE THE COEFFICIENTS OF THE QUADRATIC EQUATION FOR THE UNEQUAL
C     VARIANCES CASE
C
140   WRITE(IP,420)
      WRITE(IP,999)
      A1=DELBET**2
      A2=GNEQ*GNEQ
      A3=(RSD1*RSD1/SSQDX1)+(RSD2*RSD2/SSQDX2)
      B1=2.0D0*DELALP*DELBET
      B2=2.0D0*A2
      B3=(RSD1*RSD1*XBAR1/SSQDX1)+(RSD2*RSD2*XBAR2/SSQDX2)
      C1=DELALP**2
      C2=A2
      C3=(RSD1*RSD1/AN1)+(RSD2*RSD2/AN2)+(RSD1*RSD1*XBAR1*XBAR1/SSQDX1)+
     1(RSD2*RSD2*XBAR2*XBAR2/SSQDX2)
      A=A1-A2*A3
      B=B1+B2*B3
      C=C1-C2*C3
C     CHECK FOR ZERO COEFFICIENT OF THE QUADRATIC TERM FOR THE UNEQUAL
```

```
C     VARIANCES CASE
      IF(A.NE.0.0D0)GOTO150
      AMID=-C/B
      WRITE(IP,400)
      WRITE(IP,321)AMID
      RETURN
C
C     COMPUTE THE ROOTS OF THE QUADRATIC, AND THE MIDPOINT OF THE ROOTS
C     FOR THE UNEQUAL VARIANCES CASE
C
  150 DISC=B*B-4.0D0*A*C
C
C     CHECK FOR A NEGATIVE DISCRIMINANT FOR THE UNEQUAL VARIANCES CASE
C
      IF(DISC.GE.0.0D0)GOTO160
      WRITE(IP,410)DISC
      RETURN
  160 ROOT1=(-B+DSQRT(DISC))/(2.0D0*A)
      ROOT2=(-B-DSQRT(DISC))/(2.0D0*A)
      AMID=-B/(2.0D0*A)
      WIDTH2=ROOT1-ROOT2
      RATWID=WIDTH2/WIDTH1
      WRITE(IP,350)ROOT2
      WRITE(IP,360)ROOT1
      WRITE(IP,370)AMID
      WRITE(IP,380)WIDTH2
      WRITE(IP,999)
      WRITE(IP,999)
      WRITE(IP,390)RATWID
      RETURN
C
C
  170 FORMAT(1H ,58HSTATISTICAL ANALYSIS OF INTERSECTION OF 2 REGRESSION
     1 LINES)
  180 FORMAT(1H ,10HINPUT DATA)
  190 FORMAT(1H ,9HAN1    = ,D15.8,10X,9HAN2    = ,D15.8)
  200 FORMAT(1H ,9HXBAR1  = ,D15.8,10X,9HXBAR2  = ,D15.8)
  210 FORMAT(1H ,9HSSQDX1 = ,D15.8,10X,9HSSQDX2 = ,D15.8)
  220 FORMAT(1H ,9HALP1   = ,D15.8,10X,9HALP2   = ,D15.8)
  230 FORMAT(1H ,9HBET1   = ,D15.8,10X,9HBET2   = ,D15.8)
  240 FORMAT(1H ,9HRSD1   = ,D15.8,10X,9HRSD2   = ,D15.8)
  250 FORMAT(1H ,52H2-SIDED CONFIDENCE INTERVALS WITH PROBABILITY WIDTH
     1,D12.5,13H ARE PROVIDED)
  260 FORMAT(1H ,4HTHE ,D12.5,42H PERCENT POINT OF THE T DISTRIBUTION WI
     1TH ,I6,50H DEGREES OF FREEDOM (THE EQUAL VARIANCES CASE) IS ,D12.5
     1)
  270 FORMAT(1H ,4HTHE ,D12.5,42H PERCENT POINT OF THE T DISTRIBUTION WI
     1TH ,I6,52H DEGREES OF FREEDOM (THE UNEQUAL VARIANCES CASE) IS ,D12
     1.5)
  280 FORMAT(1H ,98HTHE RATIO OF THE LARGER ESTIMATED RESIDUAL VARIANCE
     1TO THE SMALLER ESTIMATED RESIDUAL VARIANCE IS ,D15.8)
  290 FORMAT(1H ,35HTHE MAXIMUM LIKELIHOOD ESTIMATOR = ,D15.8)
  300 FORMAT(1H ,4HS = ,D15.8)
  310 FORMAT(1H ,79HCOMPARE THE ABOVE RATIO VALUE TO THE PERCENT POINTS
     1OF THE F DISTRIBUTION WITH ,I6,5H AND ,I6,19H DEGREES OF FREEDOM)
  320 FORMAT(1H ,86H*****ERROR--THE COEFFICIENT OF THE QUADRATIC TERM IS
     1 ZERO FOR THE EQUAL VARIANCES CASE)
  321 FORMAT(1H ,57H            THE CONFIDENCE INTERVAL HAS ZERO WIDTH A
     1BOUT ,D15.8)
  330 FORMAT(1H ,97H*****ERROR--NEGATIVE DISCRIMINANT IN THE EQUAL VARIA
     1NCES CASE.  THE VALUE OF THE DISCRIMINANT IS ,D15.8)
```

```
 340 FORMAT(1H ,20HEQUAL VARIANCES CASE)
 350 FORMAT(1H ,37HLOWER LIMIT OF CONFIDENCE INTERVAL = ,D15.8)
 360 FORMAT(1H ,37HUPPER LIMIT OF CONFIDENCE INTERVAL = ,D15.8)
 370 FORMAT(1H ,37HMIDPOINT OF CONFIDENCE INTERVAL    = ,D15.8)
 380 FORMAT(1H ,37HWIDTH OF CONFIDENCE INTERVAL       = ,D15.8)
 390 FORMAT(1H ,46HTHE RATIO OF THE CONFIDENCE INTERVAL WIDTHS = ,D15.8
    1)
 400 FORMAT(1H ,88H*****ERROR--THE COEFFICIENT OF THE QUADRATIC TERM IS
    1 ZERO FOR THE UNEQUAL VARIANCES CASE)
 410 FORMAT(1H ,99H*****ERROR--NEGATIVE DISCRIMINANT IN THE UNEQUAL VAR
    1IANCES CASE.  THE VALUE OF THE DISCRIMINANT IS ,D15.8)
 420 FORMAT(1H,22HUNEQUAL VARIANCES CASE)
 998 FORMAT(1H1)
 999 FORMAT(1H0)
     END
```

```
STATISTICAL ANALYSIS OF INTERSECTION OF 2 REGRESSION LINES

INPUT DATA

AN1    =    .14000000+002          AN2    =    .11000000+002
XBAR1  =    .67500000+002          XBAR2  =   -.50000000+001
SSQDX1 =    .56875000+004          SSQDX2 =    .27500000+004
ALP1   =    .82368597+000          ALP2   =    .83584345+000
BET1   =    .62355168-003          BET2   =    .23469094-003
RSD1   =    .11459900-003          RSD2   =    .49539027-004

2-SIDED CONFIDENCE INTERVALS WITH PROBABILITY WIDTH    .95000+000 ARE PROVIDED

THE   .97500+002 PERCENT POINT OF THE T DISTRIBUTION WITH    21 DEGREES OF
      FREEDOM (THE EQUAL VARIANCES CASE) IS   .20796+001
THE   .97500+002 PERCENT POINT OF THE T DISTRIBUTION WITH    17 DEGREES OF
      FREEDOM (THE UNEQUAL VARIANCES CASE) IS   .21098+001

THE RATIO OF THE LARGER ESTIMATED RESIDUAL VARIANCE TO THE SMALLER ESTIMATED
      RESIDUAL VARIANCE IS   .5351393+001
COMPARE THE ABOVE RATIO VALUE TO THE PERCENT POINTS OF THE F DISTRIBUTION
      WITH   .12000+002 AND   .90000+001 DEGREES OF FREEDOM

THE MAXIMUM LIKELIHOOD ESTIMATOR =    .31264354+002

S =    .92500246-004
```


479-206 O - 73 - 7

```
EQUAL VARIANCES CASE

LOWER LIMIT OF CONFIDENCE INTERVAL =    .30804202+002
UPPER LIMIT OF CONFIDENCE INTERVAL =    .31727843+002
MIDPOINT OF CONFIDENCE INTERVAL    =    .31266023+002
WIDTH OF CONFIDENCE INTERVAL       =    .92364138+000


UNEQUAL VARIANCES CASE

LOWER LIMIT OF CONFIDENCE INTERVAL =    .30865357+002
UPPER LIMIT OF CONFIDENCE INTERVAL =    .31660331+002
MIDPOINT OF CONFIDENCE INTERVAL    =    .31262844+002
WIDTH OF CONFIDENCE INTERVAL       =    .79497346+000


THE RATIO OF THE CONFIDENCE INTERVAL WIDTHS =   .86069493+000
```

5. References

[1] Fisher, R. A., Statistical Methods for Research Workers, 11th edition, 144–146 (Oliver and Boyd, 1950).
[2] Kastenbaum, M. A., A confidence interval on the abscissa of the point of intersection of two fitted linear regressions, Biometrics **15**, 323–324 (1959).
[3] Hinkley, D. V., Inference about the intersection in two-phase regression, Biometrika **56**, 495–504 (1969).
[4] Robison, D. E., Estimate for the points of intersection of two polynomial regressions, J. Am. Statistical Asn. **59**, 214–224 (1964).
[5] Hudson, D. J., Fitting segmented curves whose join points have to be estimated, J. Amer. Stat. Assoc. **61**, 1097–1129 (1966).
[6] Bennett, Carl A. and Franklin, Norman L., Statistical Analysis in Chemistry and the Chemistry Industry (John Wiley & Sons, New York, 1954). pp 176–180.
[7] Aspin, A. A., An examination and further development of a formula arising in the problem of comparing two mean values, Biometrika **35**, 88 (1948).
[8] Aspin, A. A., Tables for use in comparisons whose accuracy involves two variances, separately estimated, Biometrika **36**, 290 (1949).
[9] Welch, B. L., The generalisation of "Student's" problem when several different population variances are involved, Biometrika **34**, 28 (1947).
[10] McKinney, J. E., and Belcher, H. V., to be published.

(Paper 76B3&4–370)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Volume 76B Nos. 3 and 4, July–December 1972

Weighting Methods for Monte Carlo Calculation of Polymer Configurations

Frank L. McCrackin

Institute for Materials Research, National Bureau of Standards, Washington, D.C. 20234

(September 21, 1972)

In the Rosenbluth and Rosenbluth method of computing polymer configurations, the configurations are weighted in order to remove bias of the estimated parameters of the configurations. This weighting method is investigated and generalized for importance sampling and Boltzmann factors. The estimates are found to be unbiased in the limit for an infinite sample of configurations, but to have a bias for a finite sample. The standard deviations of the estimates are also derived.

Key words: Boltzmann factors; importance sampling; Monte Carlo; Nonself-intersecting walks; polymer configurations; random walks.

Polymer molecules in solution have been simulated by non-self-intersecting random walks on a lattice by many investigators [1–7][1]. The procedure is to calculate many non-self-intersecting random walks of a given number of steps or segments, n, and calculate a parameter of each walk, such as the square of the end-to-end distance, r^2. The values of the parameter are then averaged. If the generated walks are a random sample of all non-self-intersecting walks of length n, then the average value of the parameter, such as $\langle r^2 \rangle$, will be an estimate of the average value of the parameter over all non-self-intersecting random walks of length n. Of course, for very small values of n, all possible walks may be generated and the average value of a parameter of the walks directly calculated. However, for large values of n, the number of possible walks becomes too large to generate even on a computer, so it is possible to generate only a sample of the possible walks.

The direct method of generating a sample of non-self-intersecting random walks is to generate a sample of random walks and discard those that intersect themselves. However, for large n this method is impractical because almost all generated random walks will be self-intersecting and must be discarded. Three practical methods, chain enrichment [1], dimerization [6] and the method of Rosenbluth and Rosenbluth [2] have been used to generate non-self-intersecting random walks for Monte Carlo studies of polymer configurations. This paper investigates the accuracy and bias of estimates of the parameters of walks generated by the method of Rosenbluth and Rosenbluth. Rosenbluth and Rosenbluth gave an intuitive justification, but no proof that their method is unbiased. Formulas for the bias and variance of the estimates are given in the appendix of reference [4], but without detailed derivations. Also, importance sampling and Boltzmann factors that are used in a current Monte Carlo study [11] are not considered in [4]. This paper gives complete derivations of the bias and variance of the estimates and generalizes the derivations for importance sampling and Boltzmann factors.

Although random walks are usually generated on three dimensional lattices for large values of n, the methods of generation will be illustrated for $n=4$ on a square lattice, following Rosenbluth and Rosenbluth [2]. All 25 nonintersecting random walks are shown in figure 1.

In the Rosenbluth and Rosenbluth [2] method of generating random walks, only steps for which the walk does not intersect itself are taken. Thus, in the walk shown in figure 2, either

[1] Figures in brackets indicate the literature references at the end of this paper.

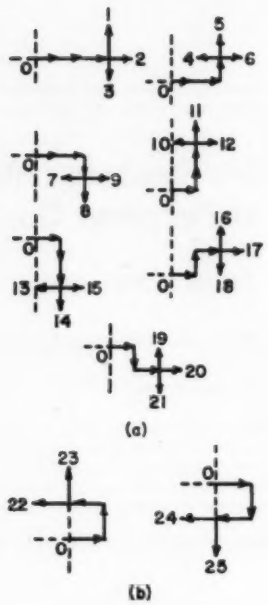


FIGURE 1. *All nonintersecting walks of 4 steps on a square lattice. The direction of the first step is fixed.*

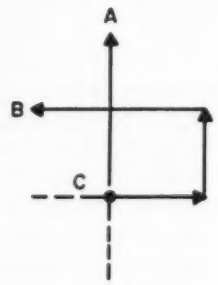


FIGURE 2. *Three possible steps for a walk on a square lattice. Step C is not allowed for a nonintersecting walk.*

step A or B would be chosen, each with probability $\frac{1}{2}$. Although there is still attrition of the walks due to trapping [3], this method allows efficient generation of long walks. However, different walks are generated with different probabilities. In the example of figure 1, walks 1 to 21 are each generated with probability $(\frac{1}{3})^3$ while walks 22 to 25 are each generated with probability of $(\frac{1}{3})^2 \frac{1}{2}$.

Consider a very large number, m, of random walks generated by this method. The average value of r^2 over the sample will be

$$\langle r^2 \rangle = \frac{1}{m} \sum_{i=1}^{m} r_i^2 \cong \sum_{j=1}^{25} r_j^2 P_j = 6.74 \tag{1}$$

where r_j^2 and P_j are the square end-to-end distance and probability for each of the 25 walks of figure 1. However, the correct average [2] obtained by averaging r^2 over the 25 configurations is 7.04. This method is seen to produce compact walks with too great a probability, so simple averages

of the walks produce incorrect, biased, results. In order to remove the bias, Rosenbluth and Rosenbluth [2] weighted each configuration to give, for any parameter v,

$$\langle \hat{v} \rangle = \frac{\sum_{k=1}^{m} v_k w_k}{\sum_{k=1}^{m} w_k} \tag{2}$$

where the weights, w_k, are the reciprocals of the probabilities of the walks. They gave an intuitive argument, but no proof, for the weighting procedure. In the above example, $w_k = 3^3$ for walks 1 to 21 and $w_k = 3^2 2$ for walks 22 to 25.

Also, the number of walks, T, of n steps was estimated by

$$\hat{T} = \frac{1}{m} \sum_{k=1}^{m} w_k. \tag{3}$$

The carets placed over v and T indicate that eqs (2) and (3) give estimates for $\langle v \rangle$ and T rather than their true values.

We first prove that eq (3) gives a correct estimate. A complication arises due to trapping [2] of the walks, when the walk cannot be continued to n steps. In this case the walk is terminated and its weighting factor is defined to be zero. Let W_T be the number of trapped configuration of walks of less than n steps, so that every walk gives either one of the T n-step walks or one of the W_T trapped configurations. The expectation [9] of w is given by the summation of $w_i P_i$ over all walks where P_i is the probability of the walk. That is:

$$E(\hat{T}) = E(w) = \sum_{i=1}^{T} w_i \frac{1}{w_i} + \sum_{i=1}^{W_T} 0 \cdot P_i = T. \tag{4}$$

Thus eq (3) estimates the number of walks. This proof follows the method of Lehman and Weiss [8].

The variance and standard deviation of T are also of interest. The variance of w is

$$\operatorname{Var} w = E(w^2) - (Ew)^2 \tag{5}$$

$$= \sum_{i=1}^{T} w_i^2 \frac{1}{w_i} - T^2$$

$$= \sum_{i=1}^{T} w_i - T^2. \tag{6}$$

Because $\hat{T}$ is an average of m independent values of w, its variance is [9]

$$\operatorname{Var} \hat{T} = \frac{1}{m} \operatorname{Var} w = \left[\sum_{i=1}^{T} w_i - T^2 \right] \Big/ m \tag{7}$$

and its standard deviation is

$$\sigma(\hat{T}) = \left[\sum_{i=1}^{T} w_i - T^2 \right]^{1/2} \Big/ \sqrt{m} \tag{8}$$

The investigation of the estimate given by eq (2) that was outlined in reference [10] is now given in detail as follows:
For a parameter v, define:

$$R=\frac{1}{m}\sum_{k=1}^{m} v_k w_k \tag{9}$$

so that

$$\langle \hat{v} \rangle = R/\hat{T}. \tag{10}$$

Let $E(R)$ and $E(\hat{T})$ be the expectation values of R and $\hat{T}$, with

$$E(\hat{T}) = T. \tag{11}$$

The expectation of R is given by the expectation of vw as

$$E(R) = \sum_{i=1}^{T} v_i w_i \frac{1}{w_i} = \sum_{i=1}^{T} v_i = T\langle v \rangle \tag{12}$$

As the sample size m tends to infinity, R and $\hat{T}$ tend to their expectation values and $\langle \hat{v} \rangle$ tends to the average of v over the walks. Thus, eq (2) is asymptotically unbiased, i.e., it is unbiased in the limit of infinite sample size.

To investigate eq (2) for large but finite values of m, $\langle \hat{v} \rangle$ is expanded about $E(R)$ and $E(\hat{T})$ to give

$$\langle \hat{v} \rangle = \frac{E(R)}{E(\hat{T})} + \frac{1}{E(\hat{T})}[R-E(R)] - \frac{E(R)}{[E(\hat{T})]^2}[\hat{T}-E(\hat{T})] - \frac{1}{[E(\hat{T})]^2}[R-E(R)][\hat{T}-E(\hat{T})]$$

$$+\frac{E(R)}{[E(\hat{T})]^3}[\hat{T}-E(\hat{T})]^2 + \text{higher order terms.} \tag{13}$$

We now take the expectation value of $\langle \hat{v} \rangle$. The first term gives $\langle v \rangle$ and the expectation value of the two first order terms vanish. Expansion and substituting for $E(R)$ and $E(\hat{T})$ from eqs (12) and (11) in eq (13) gives the approximation

$$E(\langle \hat{v} \rangle) = \langle v \rangle + [\langle v \rangle E(\hat{T}^2) - E(R\hat{T})]/T^2 \tag{14}$$

By the definition of the variance,

$$\text{Var}\,\hat{T} = E(\hat{T}^2) - [E(\hat{T})]^2. \tag{15}$$

Substituting eqs 4 and 7 gives

$$E(\hat{T}^2) = \frac{1}{m}\sum_{i=1}^{T} w_i + \frac{m-1}{m} T^2 \tag{16}$$

To evaluate $E(R\hat{T})$, we substitute from eqs (3) and (9) and rearrange terms to give

$$E(R\hat{T}) = \frac{1}{m^2} E\left(\sum_{k=1}^{m} v_k w_k \sum_{k=1}^{m} w_k\right)$$

$$= \frac{1}{m^2} E\left(\sum_{k=1}^{m} v_k w_k^2\right) + \frac{1}{m^2} E\left(\sum_{l \neq k}^{m} \sum_{l=1}^{m} v_k w_k w_l\right) \tag{17}$$

The first term on the right hand side gives $\frac{1}{m} \sum_{i=1}^{T} v_i w_i$. The second term is a summation of $m\ (m-1)$ quantities, and because $v_k w_k$ and w_l are independent, the expectation value of their product is the product of their expectation values, so

$$E\left(\sum_{l \neq k}^{m} \sum_{k=1}^{m} v_k w_k w_l\right) = m(m-1)E(v_k w_k)E(w_l)$$

$$= m(m-1)E(R)E(w) \tag{18}$$

Substituting eqs (16), (17), and (18) in eq (14) finally gives

$$E[\langle \hat{v} \rangle] = \langle v \rangle - \frac{1}{mT^2} \sum_{i=1}^{T} (v_i - \langle v \rangle) w_i \tag{19}$$

This gives the approximate bias for the average given by eq (2). That is, $\langle \hat{v} \rangle$ calculated from a sample of m walks will, on the average, differ from the true average $\langle v \rangle$ by the right-hand term. However, for increasing sample size m, the bias will approach zero, so the average is asymptotically unbiased.

For the simple example shown in figure 1, the summation was evaluated to give

$$E[\langle \hat{r}^2 \rangle] = \langle r^2 \rangle - 0.233/m$$

Thus, for a sample size of 100 walks, the mean value of r^2 calculated by eq (2) would be on the average too low by about only 0.002 lattice units squared. For longer walks on three-dimensional lattices, it is not practical to calculate the bias from eq (19). Methods of estimating the bias for these cases will be given in a later publication [11].

Using the formulas of *Ku* [12], the approximate standard deviation of $\langle \hat{v} \rangle$ (from its average value, not from the true value $\langle v \rangle$) is derived:

$$\sigma(\langle \hat{v} \rangle) = \left[\frac{1}{T^2} \sum_{i=1}^{T} (v_i - \langle v \rangle)^2 w_i\right]^{1/2} \Big/ \sqrt{m} \tag{20}$$

The standard deviation decreases as the square root of the sample size. Therefore, for sufficiently large sample size, the bias will be much smaller than the standard deviation, so the bias may be neglected. For the example of figure 1,

$$\sigma(\langle \hat{r}^2 \rangle) = 3.58/\sqrt{m}$$

For $m = 100$, the standard deviation of this estimate is about 0.36, which is much larger than the bias of the estimate.

To apply this method to long walks on various lattices, weighting factors, w, must be computed for each walk as the walk is generated. Let q be the maximum number of choices for a step of a walk on a lattice, i.e., one less than the coordination number of the lattice. Thus, $q = 3$, 5 and 11

for square, simple cubic, and face-centered cubic lattices, respectively. Also, when two nonadjacent sites of a walk lie within one lattice distance of each other, they are said to form a contact. Thus in figure 1, walks 4, 7, 18, 19, and 22 to 25 contain a contact. Now let the ith site of a walk form C_i contacts, so the next step of the walk has $q-C_i$ possible directions. One of these directions is chosen randomly, so it is chosen with a probability of $\frac{1}{q-C_i}$. Therefore, the probability of generating a particular walk of n steps is $\prod_{i=1}^{n-1} \frac{1}{q-C_i}$ and $w=\prod_{i=1}^{n-1} (q-C_i)$.

During generation of the walks, some of the walks are trapped, i.e., all surrounding sites are occupied so the walk cannot be continued. Because w_i is defined to be zero for these walks, they should not be used in the calculation of eq (2).

Average for Walks With Boltzmann Factors

In calculations for polymer walks with nearest neighbor interaction energies, the averages over parameters of the walks multiplied by Boltzmann factors $\alpha_i=\exp(-p_i\,\epsilon/kt)$ are desired, where p_i is the number of contacts of the ith walk, ϵ is the energy per contact, k is the Boltzmann constant, and t is the temperature. That is, the quantities

$$S=\sum_{i=1}^{T} \alpha_i \tag{21}$$

and

$$\langle v\rangle=\sum_{i=1}^{T} \alpha_i v_i/S \tag{22}$$

are to be estimated. The derivation follows as previously. Thus corresponding to eq (4).

$$E(w\alpha)=\sum_{i=1}^{T} w_i\alpha_i\frac{1}{w_i}=S \tag{23}$$

so an estimate of S from a sample of m walks is

$$\hat{S}=\frac{1}{m}\sum_{k=1}^{m} w_k\alpha_k \tag{24}$$

to estimate $\langle v\rangle$ we propose

$$\langle\hat{v}\rangle=\frac{\sum_{k=1}^{m} v_k\alpha_k w_k}{\sum_{k=1}^{m} \alpha_k w_k} \tag{25}$$

The derivation is the same as previously with αw replacing w and S replacing T (except for limits of the summations). Equation (19) then becomes

$$E[\langle\hat{v}\rangle]=\langle v\rangle-\frac{1}{mS^2}\sum_{i=1}^{T}(v_i-\langle v\rangle)\alpha_i w_i \tag{26}$$

and eq (20) becomes

$$\sigma(\langle\hat{v}\rangle)=\left[\frac{1}{S^2}\sum_{i=1}^{T}(v_i-\langle v\rangle)^2\,\alpha_i w_i\right]^{1/2}\Big/\sqrt{m} \tag{27}$$

Importance Sampling

In the described method, all allowable steps from a given site have been given equal probabilities. However, walks are sometimes performed in which the allowable steps are taken with different probabilities. In this way, many walks for which a desired parameter is large (important walks) will be generated. The average of the parameter over this sample of generated walks will be calculated with a higher accuracy. This technique is called Importance Sampling [13]. One method used gives twice the probability to allowable steps that point toward the origin than to other steps. Thus, in figure 3a on the rectangular lattice, step A would be taken with a probability of $\frac{1}{2}$ and steps B and C would be taken with a probability of $\frac{1}{4}$ each. In figure 3b, step A would be taken with a probability $\frac{2}{3}$ and step B with probability $\frac{1}{3}$. With this method many more coiled walks containing large number of contacts are generated than when all steps have equal probability.

For walks with large nearest-neighbor interaction energies of attraction, the Boltzmann factors α_i are large for walks containing many contacts. Then, for a parameter v such as end-to-end distance, many walks for which $v_k\alpha_k$ is large are generated by the above method so that more accurate values for the average of v by eq (25) should be obtained than for walks generated in the ordinary way. This method of generating walks will be used in reference [11].

Many other methods of choosing unequal probabilities for the steps may be used. For another example, the step in the same direction as the preceding step may be given a higher probability than the other steps. For any such method, weighting factors w_i may be calculated so that eq (2) will apply. We now derive the weighting factors for a general method of chain generation with Importance Sampling.

For any step in a walk, let there be s allowable steps. Let a multiplicity $m_k (k=1 \text{ to } s)$ be assigned to each step proportional to the probability of the step. For example, let steps toward the start of the walk be given twice the probability of other steps. Thus, in figure 3a, $s=3$, step A has a multiplicity of 2 and steps B and C each have a multiplicity of 1; in figure 3b, $s=2$, and steps A and B have multiplicities of 2 and 1 respectively. In general, the probability of a step is

$$\frac{m_c}{\sum_{k=1}^{s} m_k} \tag{28}$$

where m_c is the multiplicity of the chosen step. The probability of a given N step walk is then the product of the factors (28) for all steps of the walk and the weighting factor is the reciprocal of the probability, or the products of the factors

$$\frac{\sum_{k=1}^{s} m_k}{m_c} \tag{29}$$

over all steps of the walk.

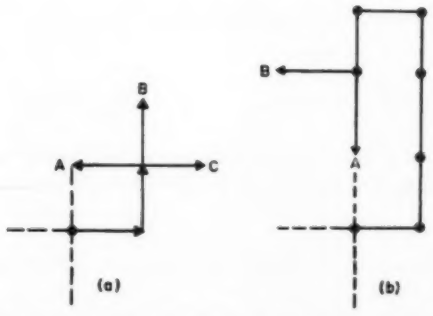


FIGURE 3. *Typical configurations on a square lattice illustrating assignment of multiplicities to allowed steps.*

Conclusions

The biases and standard deviations for parameters of walks with excluded volume generated by the method of Rosenbluth and Rosenbluth [2] have been derived and extended to the cases of importance sampling and Boltzmann factors. The direct calculation of the biases and standard deviations involve summations over all walks so is generally not feasible. However, the formulas will be used to estimate the biases and standard deviations from Monte Carlo calculations in a later paper.

I wish to thank Jacob Mazur and Charles Guttman for their help and discussions.

References

[1] Wall, F. T., Windwer, S., and Gans, P. J., in Methods in Computational Physics **1** (Academic Press, Inc., New York, 1963).
[2] Rosenbluth, M. N., and Rosenbluth, A. W., J. Chem. Phys. **23**, 356 (1955).
[3] Mazur, J., and McCrackin, F. L., J. Chem. Phys. **49**, 648 (1968).
[4] McCrackin, F., J. Chem. Phys. **47**, 1980 (1967).
[5] Medalia, A. I., in Surface and Colloid Science, Vol. 4, edited by E.Matijevic (John Wiley & Sons, 1971).
[6] Alexandrowicz, Z., J. Chem. Phys. **51**, 561 (1969).
[7] Warvari, H. E., Knaell, K. K., and Scott, R. A., III, J. Chem. Phys. **55**, 2020 (1971).
[8] Lehman, R. S., and Weiss, G. H., J. Soc. Indust. Appl. Math. **6**, 257 (1958).
[9] A discussion of calculation of expectation values is given by H. Cramer, Mathematical Methods of Statistics, (Princeton University Press, 1946).
[10] Rosenblatt, J., see Appendix of reference 4.
[11] Submitted for publication in Marcromolecules.
[12] Ku, H. H., J. Res. Nat. Bur. Stand. (U.S.), 70C (Eng. and Instr.) No. 4, 263 (Oct.–Dec. 1966). The formula of the third row of Table 1 is used.
[13] Importance Sampling is discussed for example by H. Kahn in Symposium on Monte Carlo Methods p. 149, edited by H. A. Meyer (John Wiley & Sons, Inc. New York, 1956).

(Paper 76B3&4–371)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Vol. 76B, Nos. 3 and 4, July–December 1972

Product Solutions and Separation of Variables

A. J. Goldman

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

(September 21, 1972)

This paper consists of two independent notes, whose common features are (a) their concern with "product-form" functions and (b) their use of an abstract-algebra setting to isolate the essential features of the problems treated. The first note further generalizes the generalizations given by Chu and Diaz (1965) of their observation that Euler's difference equation $y(x+1)-y(x)=f(x)$, when f is of period 1, has the product-form solution $y(x)=xf(x)$. The second note formalizes and rigorously proves the fact that a function is separable jointly in its variables iff it is separable in each individual variable.

Key words: Difference equations, functional equations, separation of variables.

1. Product Solutions of Certain Functional Equations

For this paragraph only, let $f(x)$ be a real function of period 1. Chu and Diaz,[1] apparently for the first time, observed that Euler's difference equation

$$y(x+1)-y(x)=f(x)$$

admits the simple solution $y(x)=xf(x)$. As a first generalization, they noted that the linear difference equation

$$a_n(x)y(x+n)+\ldots+a_0(x)y(x)=f(x)$$

admits the solution $y(x)=g(x)f(x)$, where $g(x)$ is any solution of

$$a_n(x)g(x+n)+\ldots+a_0(x)g(x)=1.$$

As a next generalization, they observed that the difference equation

$$F(y(x), y(x+1), \ldots, y(x+n))=f(x),$$

where F is a homogeneous function of degree 1 of its $n+1$ arguments, has the solution $y(x)=g(x)f(x)$ where $g(x)$ is any solution of

$$F(g(x), g(x+1), \ldots, g(x+n))=1.$$

In this note we seek to isolate the principle involved. Let S and S' be multiplicative semigroups, $\theta: S\to S'$ a function, and $F: S^{n+1}\to S'$ a function obeying the generalized homogeneity condition

$$F(g_0h, g_1h, \ldots, g_nh)=F(g_0, g_1, \ldots, g_n)\theta(h)$$

AMS Classification: Primary, 39A05; Secondary, 39A40.

[1] S. C. Chu and J. B. Diaz, On the Difference Equation $y(x+1)-y(x)=f(x)$, J. Math. Anal. Appl. 11 (1965), 478–481.

for all $g_0, g_1, \ldots, g_n, h \epsilon S$.

THEOREM: *Let* $T_0, T_1, \ldots, T_n$ *be homomorphisms of* S *which have a common fixed point* $f \epsilon S$. *Then the equation*

$$F(T_0 y, T_1 y, \ldots, T_n y) = c\theta(f) \tag{1}$$

admits the solution $y = gf$, *where* g *is any solution of*

$$F(T_0 g, T_1 g, \ldots, T_n g) = c \epsilon S'. \tag{2}$$

The proof is by simple calculation:

$$\begin{aligned} & F(T_0 gf, T_1 gf, \ldots, T_n gf) \\ &= F((T_0 g)(T_0 f), (T_1 g)(T_1 f), \ldots, (T_n g)(T_n f)) \\ &= F((T_0 g) f, (T_1 g) f, \ldots, (T_n g) f) \\ &= F(T_0 g, T_1 g, \ldots, T_n g)\theta(f) \\ &= c\theta(f). \end{aligned}$$

To obtain the situation discussed at the end of the first paragraph, one need only take

$S = S' =$ semigroup of real functions of a real variable ("pointwise" multiplication),

$\theta =$ identity map of S,

$c = 1 \epsilon S$,

$T =$ "step-up" operator, defined by $(Tg)(x) = g(x+1)$,

$T_i = T^i$ for $0 \leq i \leq n$.

2. On Separation of Variables

The method of separation-of-variables figures prominently in the solution of many problems of applied mathematics and mathematical physics. In this note we formalize and prove a simple principle, concerning this method, which appears to be frequently invoked on an implicit rather than explicit basis.

Let $X_1, \ldots, X_n$ be nonempty sets. We write $x = (x_1, \ldots, x_n)$ for a general member of the Cartesian product $X_1 \times \ldots \times X_n = X$, let X^i be the corresponding product with the factor X_i removed, and set $x^i = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

Let S be a commutative semigroup written multiplicatively, and $f: X \to S$ a function. We will call f *totally separable* (i.e., separable jointly in its variables) if there exist functions $f_i: X_i \to S$ for which

$$f(x) = f_1(x_1) \ldots f_n(x_n) \qquad (\text{all } x \epsilon X). \tag{3}$$

If there exist functions $g_i: X_i \to S$ and $h^i: X^i \to S$ such that

$$f(x) = g_i(x_i) h^i(x^i) \qquad (\text{all } x \epsilon X), \tag{4}$$

we call f *separable in the ith variable*. Clearly, if f is totally separable, then it is separable in each

of its variables. The "simple principle" alluded to above, is that the converse of the last statement also holds.

More specifically, we call $(g_1, \ldots, g_n)$ an *associate* of f if there exists $(h^1, \ldots, h^n)$ such that (4) holds for all i, and we will prove that under these circumstances

$$f(x) = cg_1(x_1) \ldots g_n(x_n) \qquad (\text{all } x \epsilon X) \tag{5}$$

for some $c \epsilon S$. The proof is simplest when $S = G$, a multiplicative *group*. This restriction would however rule out direct applicability to the most important case in which S is the real axis R. We could take $G = R - \{0\}$ and then perhaps handle zero values by appeal to some further "smoothness" assumption, but this is inelegant; instead we will prove the result for the case $S = G \cup \{0\}$, where 0 is the unique member of $S - G$ and is such that $s.0 = 0$ for all $s \epsilon S$. (Stripping out appropriate parts of the proof will establish the result when $S = G$.)

The proof will be by induction on n. The result is clearly true for $n = 1$, since then the function h^1 of (4) is a constant c, so that $f(x_1) = cg_1(x_1)$.

Suppose the result is true for $n = k$, that $X = X_1 \times \ldots \times X_k \times X_{k+1}$, and that $f: X \to S$ has associate $(g_1, \ldots, g_{k+1})$ with corresponding $(h^1, \ldots, h^{k+1})$ as in (4). Set $Y_i = X_i - g_i^{-1}(0)$. If any Y_i is empty, then (4) shows that $f = 0$, and so (5) holds with $c = 0$. It may therefore be assumed that each Y_i is nonempty; let $Y = Y_1 \times \ldots \times Y_{k+1}$. We will prove the existence of $c \epsilon S$ for which (5) holds for all $x \epsilon Y$; it will then, by (4), automatically hold for all $x \epsilon X - Y$ as well.

For $1 \leq i \leq k$,

$$f(x) = g_{k+1}(x_{k+1}) h^{k+1}(x^{k+1}) = g_i(x_i) h^i(x^i),$$

and so

$$[g_{k+1}(x_{k+1})]^{-1} h^i(x^i) = [g_i(x_i)]^{-1} h^{k+1}(x^{k+1}) \qquad (\text{all } x \epsilon Y).$$

The left-hand side is independent of x_i and the right-hand side is independent of x_{k+1}, so that each is equal to some $G_i(x)$ which is independent of both x_i and x_{k+1}. If we set $Y' = Y_1 \times \ldots \times Y_k$ and write $x' = (x_1, \ldots, x_k)$, then

$$f(x) = f(x'; x_{k+1}) = g_{k+1}(x_{k+1})\ g_i(x_i)\ H_i(x') \qquad (\text{all } x \epsilon Y),$$

for $1 \leq i \leq k$, where $H_i(x') = G_i(x)$ is independent of x_i. Hence for each $x_{k+1} \epsilon Y_{k+1}$, the function $F: Y' \to S$ defined by

$$F(x') = [g_{k+1}\ (x_{k+1})]^{-1}\ f(x'; x_{k+1})$$

has $(g^1, \ldots, g^k)$ as associate;[2] by the induction hypothesis, there is a $c(x_{k+1}) \epsilon S$ such that

$$F(x') = g_1\ (x_1) \ldots g_k\ (x_k)\ c(x_{k+1}) \qquad (\text{all } x' \epsilon Y')$$

and so

$$f(x) = g_1\ (x_1) \ldots g_{k+1}\ (x_{k+1}) c(x_{k+1}) \qquad (\text{all } x \epsilon Y).$$

But by (4),

$$f(x) = g_{k+1}\ (x_{k+1})\ h^{k+1}\ (x^{k+1}) \qquad (\text{all } x \epsilon Y),$$

[2] More precisely, the vector of function-restrictions $g_i | Y_i$ for $1 \leq i \leq k$.

so that

$$c(x_{k+1}) = [g_1(x_1) \ . \ . \ . \ g_k\ (x_k)]^{-1}\ h^{k+1}\ (x^{k+1}).$$

Since the left-hand side depends only on x_{k+1} while the right-hand side is independent of this variable, $c(x_{k+1})$ must be a constant, and so (5) is established for all $x \epsilon Y$, completing the proof.

(Paper 76B3&4–372)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Vol. 76B, Nos. 3 and 4, July–December 1972

A Note on the Multiplicative Property of the Smith Normal Form*

Marvin Marcus and Ernest E. Underwood*****

(September 21, 1972)

This paper contains an elementary proof of the fact that if A and B are n-square matrices over a principal ideal domain R with relatively prime determinants, then $S(AB)=S(A)S(B)$ where $S(A)$ is the Smith normal form of A.

Key words: Compound; divisors; matrix.

In [2; p. 33][1] the following interesting result appears: *If* A *and* B *are* n*-square matrices over a principal ideal domain* R *and* g.c.d. (det (A), det (B))$=1$ *then* S(AB)$=$S(A)S(B) *where* S(A) *is the Smith normal form of* A.

The purpose of this note is to present a simple proof of the result that uses elementary properties of compound matrices.

LEMMA. *Let* $Q=\operatorname{diag}(q_1, \ldots, q_n)$, $P=\operatorname{diag}(p_1, \ldots, p_n)$, $q_1 \mid q_j$, $p_1 \mid p_j$, $j=1, \ldots, n$ *and* g.c.d. $(p_i, q_j)=1$, i, j$=1, \ldots,$ n. *Let* U *be an* n*-square matrix with the property that* g.c.d. $(u_{11}, u_{21}, \ldots, u_{n1})=$ g.c.d. $(u_{11}, u_{12}, \ldots, u_{1n})=1$. *Then the* g.c.d. *of all the entries in* QUP *is* p_1q_1.

PROOF. Obviously $p_1q_1 \mid QUP$, i.e., p_1q_1 divides every entry of QUP. Write $QUP=p_1q_1D$. Suppose that $p \mid D$ where p is a prime. It is simple to see that the first row and column of D are respectively

$$D_{(1)}=[u_{11}, u_{12}p_2', \ldots, u_{1n}p_n'], \qquad p_i'=p_ip_1^{-1},$$

and

$$D^{(1)}=[u_{11}, q_2'u_{21}, \ldots, q_n'u_{n1}], \qquad q_i'=q_iq_1^{-1}.$$

Now $p \mid D_{(1)}$ and since g.c.d.$(u_{11}, \ldots, u_{1n})=1$ we conclude that $p \mid p_k'$, for some $k=2, \ldots, n$. Similarly $p \mid D^{(1)}$ so $p \mid q_l'$ for some $l=2, \ldots, n$. But then $p \mid p_k' \mid p_k$, $p \mid q_l' \mid q_l$ and this contradicts g.c.d.$(p_k, q_l)=1$.

Since A and $S(A)$ are equivalent it follows immediately that AB and $S(A)US(B)$ are equivalent where U is unimodular. Thus

$$S(AB)=S(S(A)US(B)). \tag{1}$$

Since it is obvious that $d_k(S(A)S(B))=d_k(A)d_k(B)$, $k=1, \ldots, n$, we need only show that

$$d_k(A)d_k(B)=d_k(S(A)US(B)), k=1, \ldots, n,$$

AMS Subject Classification: Primary 15A21; Secondary 15A33.
*An invited paper. The research of the first author was supported by U.S. Air Force Office of Scientific Research Grant AFOSR 72-2164, and the research of both authors was supported under the auspices of An Intermediate Short Course in Linear Algebra for College Teachers of Mathematics, N.S.F. Grant GY9414.
**Present address: Department of Mathematics, University of California, Santa Barbara, California 93106.
***Present address: Department of Mathematics, Utah State University, Logan, Utah 84321.
[1] Figures in brackets indicate the literature references at the end of this paper.

to complete the proof.

If $C_k(X)$ denotes the kth compound of the matrix X [1; p. 16] then we immediately have

$$C_k(S(A)US(B)) = C_k(S(A))C_k(U)C_k(S(B)) \tag{2}$$

and the matrix $C_k(U)$ is unimodular. For, $C_k(UV) = C_k(I_n)$ implies $C_k(U)C_k(V) = C_k(I_n)$ and thus $C_k(U)$ has an inverse over R if U does. (Or more simply, apply the Sylvester-Franke theorem to see that $C_k(U)$ has a unit determinant.) Hence the entries in the first row (column) of $C_k(U)$ are relatively prime. The divisibility properties of the determinantal divisors together with the hypothesis that g.c.d. $(d_n(A), d_n(B)) = 1$ imply that any two main diagonal elements of the diagonal matrices $C_k(S(A))$ and $C_k(S(B))$ are relatively prime. Moreover the 1,1 entry of $C_k(S(A))$ is $d_k(A)$ and similarly for $C_k(S(B))$. We can now apply the lemma to the matrix on the right in (2) to conclude that $d_k(A)d_k(B)$ is the g.c.d. of the entries in $C_k(S(A))C_k(U)C_k(S(B))$; i.e., $d_k(A)d_k(B)$ is the g.c.d. of the entries in $C_k(S(A)US(B))$. But in view of (1),

$$d_k(S(AB)) = d_k(A)d_k(B)$$

and the proof is complete.

References

[1] Marcus, Marvin, and Minc, Henryk, A Survey of Matrix Theory and Matrix Inequalities, (Allyn and Bacon, Boston, Mass., 1964).

[2] Newman, Morris, Integral Matrices, (Academic Press, New York, 1972).

(Paper 76B3&4–373)

JOURNAL OF RESEARCH of the National Bureau of Standards — B. Mathematical Sciences
Vol. 76B, Nos. 3 and 4, July-December 1972

Publications of the National Bureau of Standards*

Citations

J. Res. Nat. Bur. Stand. (U.S.), **76A** *(Phys. & Chem.),* No. 4, (July-August 1972), SD Catalog No. C13.22/sec.A:76/4.

Hamer, W. J., Skapars, A., and Field, B. F., **The emf-temperature coefficient of "acid" standard cells of the saturated cadmium sulfate type from 15 to 40 °C.**

Rebbert, R. E., and Ausloos, P., **Pulse radiolysis of neopentane in the gas phase.**

Haller, W., and Simmons, J. H., **Anatomy and thermal history of laser self-focusing damage tracks in glass.**

Fussell, W. B., **Normal emissivity of an isothermal, diffusely reflecting cylindrical cavity (with top) as a function of inside radius.**

Fanconi, B. M., **Lattice vibrations of antiparallel chain sheet structures.**

Monogr. 124. **Reference tables for low-temperature thermocouples,** L. L. Sparks, R. L. Powell, and W. J. Hall, Nat. Bur. Stand. (U.S.), Monogr. 124, 61 pages (June 1972) 60 cents, SD Catalog No. C13.44:124.

Monogr. 127. **NBS papers on underground corrosion of steel piling, 1962-1971,** W. J. Schwerdtfeger and M. Romanoff, Nat. Bur. Stand. (U.S.), Monogr. 127, 63 pages (Mar. 1972) 65 cents, SD Catalog No. C13.44:127.

H105-1. (Revised 1972). **Specifications and tolerances for reference standards and field standard weights and measures. 1. Specifications and tolerances for field standard weights (NBS class F),** T. M. Stabler, Nat. Bur. Stand. (U.S.), Handb. 105-1, 12 pages (July 1972) 25 cents, SD Catalog No. C13.11:105-1.

H110-1. **National Electrical Safety Code. Part 1. Rules for the installation and maintenance of electrical supply stations and equipment,** Nat. Bur. Stand. (U.S.), Handb. 110-1, 43 pages (June 1972) 30 cents, SD Catalog No. C13.11:110-1.

H111. **American National Standard. Radiation safety for x-ray diffraction and fluorescence analysis equipment,** Nat. Bur. Stand. (U.S.), Handb. 111, 20 pages (June 1972) 30 cents, SD Catalog No. C13.11:111.

SP260-33. **Standard reference materials: Comparison of original and supplemental SRM 705, narrow molecular weight distribution polystyrene,** H. L. Wagner, Nat. Bur. Stand. (U.S.), Spec. Publ. 260-33, 30 pages (May 1972) 35 cents, SD Catalog No. C13.10:260-33.

SP260-34. **Standard reference materials: Thermoelectric voltage of silver-28 atomic percent gold thermocouple wire, SRM 733, versus common thermocouple materials (between liquid helium and ice fixed points),** L. L. Sparks and J. G. Hust, Nat. Bur. Stand. (U.S.), Spec. Publ. 260-34, 34 pages (Apr. 1972) 40 cents, SD Catalog No. C13.10:260-34.

SP260-35. **Standard reference materials: Thermal conductivity of austenitic stainless steel, SRM 735, from 5 to 280 K,** J. G. Hust and L. L. Sparks, Nat. Bur. Stand. (U.S.), Spec. Publ. 260-35, 22 pages (Apr. 1972) 35 cents, SD Catalog No. C13.10:260-35.

SP260-36. **Standard reference materials: A referee method for the determination of calcium in serum,** J. P. Cali, J. Mandel, and L. Moore, Nat. Bur. Stand. (U.S.), Spec. Publ. 260-36, 136 pages (May 1972) $1.25, SD Catalog No. C13.10:260-36.

SP260-37. **Standard reference materials: Methods of analysis of NBS clay standards,** J. I. Shultz, R. K. Bell, T. C. Rains, and O. Menis, Nat. Bur. Stand. (U.S.), Spec. Publ. 260-37, 86 pages (June 1972) 75 cents, SD Catalog No. C13.10:260-37.

SP260-38. **Standard reference materials: Preparation and calibration of standards of spectral specular reflectance,** J. C. Richmond and J. J. Hsia, Nat. Bur. Stand. (U.S.), Spec. Publ. 260-38, 57 pages (May 1972) 60 cents, SD Catalog No. C13.10:260-38.

SP260-39. **Standard reference materials: The eddy current decay method for resistivity characterization of high purity metals,** A. F. Clark, V. A. Deason, J. G. Hust, and R. L. Powell, Nat. Bur. Stand. (U.S.), Spec. Publ. 260-39, 53 pages (May 1972) 55 cents, SD Catalog No. C13.10:260-39.

SP260-40. **Standard reference materials: Selection of differential thermal analysis temperature standards through a cooperative study (SRM 758, 759, 760),** H. G. McAdie, P. D. Garn, and O. Menis, Nat. Bur. Stand. (U.S.), Spec. Publ. 260-40, 71 pages (Aug. 1972) 65 cents, SD Catalog No. C13.10:260-40.

SP260-41. **Standard reference materials: Use of standard light-sensitive paper for calibrating carbon arcs used in testing textiles for colorfastness to light,** L. A. Wood and P. J. Shouse, Nat. Bur. Stand. (U.S.), Spec. Publ. 260-41, 24 pages (Aug. 1972) 30 cents, SD Catalog No. C13.10:260-41.

SP300. Vol. 5. **Precision measurement and calibration. Selected NBS papers on frequency and time,** B. E. Blair and A. H. Morgan, Editors, Nat. Bur. Stand. (U.S.), Spec. Publ. 300, Vol. 5, 565 pages (June 1972) $6.00, SD Catalog No. C13.10:300, Vol. 5.

SP300. Vol. 9. **Precision measurement and calibration. Selected NBS papers on colorimetry,** I. Nimeroff, Editor, Nat. Bur. Stand. (U.S.), Spec. Publ. 300, Vol. 9, 460 pages (June 1972) $5.50, SD Catalog No. C13.10:300, Vol. 9.

SP305. Supplement 3. **Publications of the National Bureau of Standards 1971 catalog. A compilation of abstracts and key word and author indexes,** B. L. Oberholtzer, Nat. Bur. Stand. (U.S.), Spec. Publ. 305, Suppl. 3, 342 pages (July 1972) $3.00, SD Catalog No. C13.10:305, Suppl. 3.

SP330. 1972 Edition. **The international system of units (SI),** C. H. Page and P. Vigoureux, Editors, Nat. Bur. Stand. (U.S.), Spec. Publ. 330, 45 pages (Apr. 1972) 30 cents, SD Catalog No.

479-206 O - 73 - 8

C13.10:330/2. Supersedes NBS Special Publication 330, 1971 Edition.

SP348. **The science of ceramic machining and surface finishing.** Proceedings of a symposium sponsored by the American Ceramic Society, the Office of Naval Research, and the National Bureau of Standards, Nov. 2-4, 1970, held at NBS, Gaithersburg, Md., S. J. Schneider, Jr., and R. W. Rice, Editors, and C. F. Bersch, A. M. Diness, and J. B. Wachtman, Jr., Associate Editors, Nat. Bur. Stand. (U.S.), Spec. Publ. 348, 431 pages (May 1972) $5.25, SD Catalog No. C13.10:348.

SP349. **Heavy-atom kinetic isotope effects, an indexed bibliography,** M. J. Stern and M. Wolfsberg, Nat. Bur. Stand. (U.S.), Spec. Publ. 349, 39 pages (June 1972) 45 cents, SD Catalog No. C13.10:349.

SP351. **Analytical chemistry: Key to progress on national problems.** Proceedings of the 24th Annual Summer Symposium on Analytical Chemistry sponsored by *Analytical Chemistry*, ACS Division of Analytical Chemistry, and NBS Analytical Chemistry Division, and held at the National Bureau of Standards, Gaithersburg, Md., June 16-18, 1971, W. W. Meinke and J. K. Taylor, Editors, Nat. Bur. Stand. (U.S.), Spec. Publ. 351, 481 pages (Aug. 1972) $3.50, SD Catalog No. C13.10:351.

SP358. **56th national conference on weights and measures 1971,** F. C. Bell and H. F. Wollin, Report Editors, Nat. Bur. Stand. (U.S.), Spec. Publ. 358, 252 pages (Mar. 1972) $1.50, SD Catalog No. C13.10:358.

SP361. Volume 2. **Performance concept in buildings. Opening addresses, rapporteur reviews, and discussions,** Proceedings of a symposium jointly sponsored by the International Union of Testing and Research Laboratories for Materials and Structures (RILEM), the American Society for Testing and Materials (ASTM), and the International Council for Building Research Studies and Documentation (CIB), Philadelphia, Pa., May 2-5, 1972, B. E. Foster, Editor, Nat. Bur. Stand. (U.S.), Spec. Publ. 361, Vol. 2, 169 pages (Sept. 1972) $1.75, SD Catalog No. C13.10:361, Vol. 2.

SP362. **Chemical kinetics in the C-O-S and H-N-O-S systems: A bibliography—1899 through June 1971,** F. Westley, Nat. Bur. Stand. (U.S.), Spec. Publ. 362, 68 pages (Apr. 1972) 70 cents, SD Catalog No. C13.10:362.

SP363. **Bibliography on atomic energy levels and spectra, July 1968 through June 1971,** L. Hagan and W. C. Martin, Nat. Bur. Stand. (U.S.), Spec. Publ. 363, 103 pages (June 1972) $1.00, SD Catalog No. C13.10:363.

SP364. **Solid state chemistry.** Proceedings of the 5th Materials Research Symposium sponsored by the Institute for Materials Research, National Bureau of Standards, Oct. 18-21, 1971, held at Gaithersburg, Md., R. S. Roth and S. J. Schneider, Jr., Editors, Nat. Bur. Stand. (U.S.), Spec. Publ. 364, 799 pages (July 1972) $7.50, SD Catalog No. C13.10:364.

SP365. **Metric conversion card,** J. V. Odom, Nat. Bur. Stand. (U.S.), Spec. Publ. 365, 2 pages (July 1972) 10 cents ($6.25 for 100 copies), SD Catalog No. C13.10:365.

AMS62. **Cyclic designs,** J. A. John, F. W. Wolock, H. A. David, J. M. Cameron, and J. A. Speckman, Nat. Bur. Stand. (U.S.), Appl. Math. Ser. 62, 79 pages (Apr. 1972) 75 cents, SD Catalog No. C13.32:62.

NSRDS-NBS39. **Tables of molecular vibrational frequencies, consolidated Volume I,** T. Shimanouchi, Nat. Stand. Ref. Data Ser., Nat. Bur. Stand. (U.S.), 39, 164 pages (June 1972) $3.00, SD Catalog No. C13.48:39.

NSRDS-NBS40. **A multiplet table of astrophysical interest, (Revised Edition), Part I—Table of multiplets, and Part II— Finding list of all lines in the table of multiplets,** C. E. Moore, Nat. Stand. Ref. Data Ser., Nat. Bur. Stand. (U.S.), 40, 253 pages (Feb. 1972) $2.00, SD Catalog No. C13.48:40.

NSRDS-NBS41. **Crystal structure transformations in binary halides,** C. N. R. Rao and M. Natarajan, Nat. Stand Ref. Data Ser., Nat. Bur. Stand. (U.S.), 41, 53 pages (July 1972) 55 cents, SD Catalog No. C13.48:41.

BSS41. **Performance of a single-stack DWV system utilizing low-angle stack-branch confluence and bottom shunt venting,** R. S. Wyly and G. C. Sherlin, Nat. Bur. Stand. (U.S.), Bldg. Sci. Ser. 41, 24 pages (Apr. 1972) 35 cents, SD Catalog No. C13.29/2:41.

BSS42. **Design loads for inserts embedded in concrete,** T. W. Reichard, E. F. Carpenter, and E. V. Leyendecker, Nat. Bur. Stand. (U.S.), Bldg. Sci. Ser. 42, 28 pages (May 1972) 35 cents, SD Catalog No. C13.29/2:42.

BSS43. **Paper honeycomb sandwich panels as lightweight structural components,** T. W. Reichard, Nat. Bur. Stand. (U.S.), Bldg. Sci. Ser. 43, 7 pages (Apr. 1972) 25 cents, SD Catalog No. C13.29/2:43.

FIPS PUB 12-1. **Federal information processing standards index,** H. E. McEwen, Standards Coordinator, Nat. Bur. Stand. (U.S.), Fed. Info. Process. Stand. Publ. (FIPS Pub.) 12-1, 140 pages (1972) $1.25, SD Catalog No. C13.52:12-1.

PS52-71. **Polytetrafluoroethylene (PTFE) plastic tubing,** L. H. Breden, Technical Standards Coordinator, Nat. Bur. Stand. (U.S.), Prod. Stand. 52-71, 12 pages (Apr. 1971) 20 cents, SD Catalog No. C13.20/2:52-71.

PS53-72. **Glass-fiber reinforced polyester structural plastic panels,** L. H. Breden, Technical Standards Coordinator, Nat. Bur. Stand. (U.S.), Prod. Stand. 53-72, 13 pages (Apr. 1972) 20 cents, SD Catalog No. C13.20/2:53-72.

TN379-1. **Standard time and frequency: Its generation, control, and dissemination from the National Bureau of Standards Time and Frequency Division,** J. B. Milton, Nat. Bur. Stand. (U.S.), Tech. Note 379-1, 26 pages (June 1972) 35 cents, SD Catalog No. C13.46:379-1.

TN467. **Activation analysis: A bibliography through 1971,** G. J. Lutz, R. J. Boreni, R. S. Maddock, and J. Wing, Editors, Nat. Bur. Stand. (U.S.), Tech. Note 467, 892 pages (Aug. 1972) $7.00, SD Catalog No. C13.46:467.

TN587. **Organic Chemistry Section: Summary of activities, July 1970 to June 1971,** R. Schaffer, Editor, Nat. Bur. Stand. (U.S.), Tech. Note 587, 131 pages (Aug. 1972) $1.25, SD Catalog No. C13.46:587.

TN594-1. **Optical radiation measurements: Fundamental principles of absolute radiometry and the philosophy of this NBS program (1968 to 1971),** J. Geist, Nat. Bur. Stand. (U.S.), Tech. Note 594-1, 59 pages (June 1972) 60 cents, SD Catalog No. C13.46:594-1.

TN614. **Longitudinal magnetoresistance in polar semiconductors,** R. L. Peterson, Nat. Bur. Stand. (U.S.), Tech. Note 614, 47 pages (Apr. 1972) 50 cents, SD Catalog No. C13.46:614.

TN615. **WR15 thermal noise standard,** W. C. Daywitt, W. J. Foote, and E. Campbell, Nat. Bur. Stand. (U.S.), Tech. Note 615, 154 pages (Mar. 1972) $1.25, SD Catalog No. C13.46:615.

TN616. **Frequency standards and clocks: A tutorial introduc-**

tion, H. Hellwig, Nat. Bur. Stand. (U.S.), Tech. Note 616, 69 pages (Apr. 1972) 70 cents, SD Catalog No. C13.46:616.

TN617. **Thermophysical properties of parahydrogen from the freezing liquid line to 5000 R for pressures to 10,000 psia,** R. D. McCarty and L. A. Weber, Nat. Bur. Stand. (U.S.), Tech. Note 617, 169 pages (Apr. 1972) $1.50, SD Catalog No. C13.46:617.

TN618. **WR15 microwave calorimeter and bolometer unit,** M. E. Harvey, Nat. Bur. Stand. (U.S.), Tech. Note 618, 41 pages (May 1972) 75 cents, SD Catalog No. C13.46:618.

TN710-2. **Building research translation: An investigation of the protection of dwellings from external noise through facade walls,** P. Gilbert, Nat. Bur. Stand. (U.S.), Tech. Note 710-2, 26 pages (Mar. 1972) 35 cents, SD Catalog No. C13.46:710-2.

TN710-3. **Building research translation: New regulation on ventilation of dwellings, fixed heating facilities, and flues,** A. Chalandon, M. Schumann, and P. Dechartre, Nat. Bur. Stand. (U.S.), Tech. Note 710-3, 46 pages (May 1972) 50 cents, SD Catalog No. C13.46:710-3.

TN710-4. **Building research translation: Thermal comfort requirement adjacent to cold walls—application to glazed opening,** J. Anquez and M. Croiset, Nat. Bur. Stand. (U.S.), Tech. Note 710-4, 54 pages (May 1972) 55 cents, SD Catalog No. C13.46:710-4.

TN710-5. **Building research translation: Use of an air-to-air heat exchanger to recover heat from air exhausted by mechanical ventilation,** P. Garriver, Nat. Bur. Stand. (U.S.), Tech. Note 710-5, 62 pages (July 1972) 60 cents, SD Catalog No. 13.46:710-5.

TN715. **Tabulation of published data on electron devices of the U.S.S.R. through December 1971,** C. P. Marsden, Nat. Bur. Stand. (U.S.), Tech. Note 715, 108 pages (June 1972) $1.00, SD Catalog No. C13.46:715.

TN716. **The ideal Lovibond color system for CIE standard illuminants A and C shown in three colorimetric systems,** G. W. Haupt, J. C. Schleter, and K. L. Eckerle, Nat. Bur. Stand. (U.S.), Tech. Note 716, 115 pages (Apr. 1972) $1.00, SD Catalog No. C13.46:716.

TN717. **Methods of measurement for semiconductor materials, process control, and devices. Quarterly Report, July 1 to September 30, 1971,** W. M. Bullis, Editor, Nat. Bur. Stand. (U.S.), Tech. Note 717, 52 pages (Apr. 1972) 55 cents, SD Catalog No. C13.46:717.

TN718. **Magnetic recording of acoustic data on audiofrequency tape recorders,** E. D. Burnett, E. L. R. Corliss, and R. D. Berendt, Nat. Bur. Stand. (U.S.), Tech. Note 718, 29 pages (Apr. 1972) 40 cents, SD Catalog No. C13.46:718.

TN719. **A simple correction procedure for quantitative electron probe microanalysis,** K. F. J. Heinrich, R. L. Myklebust, H. Yakowitz, and S. D. Rasberry, Nat. Bur. Stand. (U.S.), Tech. Note 719, 49 pages (May 1972) 55 cents, SD Catalog No. C13.46:719.

TN720. **A simple hydraulic sinusoidal pressure calibrator,** J. S. Hilten, P. S. Lederer, and J. Sethian, Nat. Bur. Stand. (U.S.), Tech. Note 720, 27 pages (Apr. 1972) 35 cents, SD Catalog No. C13.46:720.

TN721. **NBS Special Foreign Currency Program in Israel 1970-71,** H. S. Peiser and M. Klein, Nat. Bur. Stand. (U.S.), Tech. Note 721, 175 pages (Apr. 1972) $1.50, SD Catalog No. C13.46:721.

TN722. **Tensile creep of boron epoxy and boron epoxy-reinforced 7075-T6 aluminum alloy,** D. J. Chwirut, Nat. Bur. Stand. (U.S.), Tech. Note 722, 60 pages (May 1972) 60 cents, SD Catalog No. C13.46:722.

TN723. **A laser technique for investigating the effects of thermal transients on pressure transducer performance characteristics,** P. S. Lederer and J. S. Hilten, Nat. Bur. Stand. (U.S.), Tech. Note 723, 43 pages (May 1972) 45 cents, SD Catalog No. C13.46:723.

TN724. **Properties of selected superconductive materials,** B. W. Roberts, Nat. Bur. Stand. (U.S.), Tech. Note 724, 100 pages (June 1972) $1.00, SD Catalog No. C13.46:724.

TN725. **A neutron moments computer code, moment I,** C. M. Eisenhauer, G. L. Simmons, and L. V. Spencer, Nat. Bur. Stand. (U.S.), Tech. Note 725, 38 pages (May 1972) 50 cents, SD Catalog No. C13.46:725.

TN726. **Testing and fabrication of wire-bond electrical connections—a comprehensive survey,** H. A. Schafft, Nat. Bur. Stand. (U.S.), Tech. Note 726, 140 pages (Sept. 1972) $1.25, SD Catalog No. C13.46:726.

TN727. **Methods of measurement for semiconductor materials, process control, and devices. Quarterly report October 1 to December 31, 1971,** W. M. Bullis, Editor, Nat. Bur. Stand. (U.S.), Tech. Note 727, 79 pages (June 1972) 70 cents, SD Catalog No. C13.46:727.

TN728. **Torsional instabilities in composite and composite-reinforced aluminum-alloy thin-walled cylinders,** D. E. Marlowe, Nat. Bur. Stand. (U.S.), Tech. Note 728, 77 pages (June 1972) 70 cents, SD Catalog No. C13.46:728.

TN729. **Design, construction, and testing of a new high accuracy spectrophotometer,** K. D. Mielenz and K. L. Eckerle, Nat. Bur. Stand. (U.S.), Tech. Note 729, 60 pages (June 1972) 60 cents, SD Catalog No. C13.46:729.

TN730. **Manual and automated fingerprint registration,** J. H. Wegstein, Nat. Bur. Stand. (U.S.), Tech. Note 730, 26 pages (June 1972) 35 cents, SD Catalog No. C13.46:730.

TN731. **Calibration of secondary standard magnetic tape cassettes (computer amplitude reference) Phase I,** S. B. Geller and P. A. Mantek, Nat. Bur. Stand. (U.S.), Tech. Note 731, 60 pages (July 1972) 60 cents, SD Catalog No. C13.46:731.

TN732. **A computer terminal network for transparent stimulation of the user of an on-line retrieval system,** S. Treu, Nat. Bur. Stand. (U.S.), Tech. Note 732, 39 pages (July 1972) 45 cents, SD Catalog No. C13.46:732.

TN733. **Methods of measurement for semiconductor materials, process control, and devices. Quarterly report January 1 to March 31, 1972,** W. M. Bullis, Editor, Nat. Bur. Stand. (U.S.), Tech. Note 733, 59 pages (Sept. 1972) 60 cents, SD Catalog No. C13.46:733.

TN734. **OMNITAB II. Segmentation structure for the SCOPE operating system,** S. T. Peavy and R. N. Varner, Nat. Bur. Stand. (U.S.), Tech. Note 734, 25 pages (June 1972) 35 cents, SD Catalog No. C13.46:734.

TN738. **Subroutine for the calculation of CODEN check characters,** D. Garvin, Nat. Bur. Stand. (U.S.), Tech. Note 738, 12 pages (Sept. 1972) 25 cents, SD Catalog No. C13.46:738.

The following Bibliography Series are available by purchase from the National Technical Information Service (NTIS), Springfield, Va. 22151.

COM-72-10014. Kondratiev, V. N., **Rate constants of gas phase reactions,** translated by L. J. Holtschlag and edited by R. M. Fristom, 433 pages, January 1972. (Johns Hopkins University, Applied Physics Laboratory, Silver Spring, Md.) $8.50.

COM-72-10266. **Index and cumulative list of papers on radiation chemistry,** Vol. IV, Nos. 27-52, July through December 1971, 273 pages, January 1972 (Radiation Chemistry Data Center, Radiation Laboratory, University of Notre Dame) $6.00.

The following are outside publications by the NBS staff, as soon after issuance as practical. For completeness, earlier references not previously reported may be included from time to time. The five-digit number at the beginning of each entry is used for record keeping only and should not be used when requesting reprints or other information.

12688. Carpenter, B. S., **Quantitative applications of the nuclear track technique**(Proc. Inter/Micro 71 Conf., London, England, Sept. 17-24, 1971), *Microscope* **20,** 175-182 (1972).

12689. Carpenter, B. S., LaFleur, P. D., **Observing proton tracks in cellulose nitrate,** *Int. J. Appl. Radiat. Isotop.* **23,** 157-159 (1972).

12690. Clough, R. B., **Mobile dislocation density and velocity in plastically deformed aluminum at room temperature,** *Scr. Met.* **6,** No. 4, 293-298 (1972).

12691. Codling, K., Madden, R. P., **Resonances in the photoionization continuum of Kr and Xe,** *Phys. Rev. A* **4,** No. 6, 2261-2263 (Dec. 1971).

12692. Cooper, J. W., LaVilla, R. E., **"Semi-Auger" processes in L_{23} emission in Ar and KCl,** *Phys. Rev. Lett.* **25,** No. 26, 1745-1748 (Dec. 28, 1970).

12693. Coxon, B., **Application of internuclear, double-resonance techniques to carbohydrates. Detection of small coupling-constants,** *Carbohydrate Res.* **18,** No. 3, 427-442 (1971).

12694. Dodge, W. R., Murphy, J. J., II, **Ratio of the ^{4}He(γ,p) and ^{4}He(γ,n) cross sections,** *Phys. Rev. Lett.* **28,** No. 13, 839-843 (Mar. 27, 1972).

12695. Durst, R. A., Staples, B. R., Paabo, M., **Activity standards for ion-selective electrodes,** *Experientia Suppl.* **18,** 275-279 (1971).

12696. Eisen, H., Rosenstein, M., Silverman, J., **2·00-MeV electron depth-dose measurements in aluminum, copper and tin absorbers using a radiochromic dye film,** *Int. J. Appl. Radiat. Isotop.* **23,** 97-108 (Apr. 1972).

12697. Fulmer, C. B., Toth, K. S., Williams, I. R., Handley, T. H., Dell, G. F., Callis, E. L., Jenkins, T. M., Wyckoff, J. M., **Photonuclear reactions in iron and aluminum bombarded with high-energy electrons,** *Phys. Rev. C* **2,** No. 4, 1371-1378 (Oct. 1970).

12698. Gebbie, H. A., Bohlander, R. A., **Nonresonant cavity as a long path absorption cell,** *Appl. Opt.* **11,** No. 4, 723-728 (Apr. 1972).

12699. Geltman, S., **A high energy approximation: I. Proton-hydrogen charge transfer,** *J. Phys. B: At. Mol. Phys.* **4,** No. 10, 1288-1298 (Oct. 1971).

12700. Hayward, E., Gibson, B. F., O'Connell, J. S., **Consequences of isospin sum rules for photonuclear reactions,** *Phys. Rev. C* **5,** No. 3, 846-852 (Mar. 1972).

12701. MacDonald, R. A., **Neutron scattering due to lattice distortion around point defects,** *J. Phys. F: Metal Phys.* **2,** 209-218 (Mar. 1972).

12702. Moore, G. A., **Recent progress in automatic image analysis,** *J. Micros.* **95,** Pt. 1, 105-118 (Feb. 1972).

12703. Oneal, G., Jr., Harris, W. P., **Three-terminal cell for thin film dielectric measurements** (Proc. Electrical Insulation and Dielectric Phenomena, Buck Hill Falls, Pa., Oct. 20-22, 1969), Chapter in *1969 Annual Report of the Conference on Electrical Insulation and Dielectric Phenomena,* pp. 164-170 (National Research Council, National Academy of Sciences, Washington, D.C., 1970).

12704. Newman, M., **2-Generator groups and parabolic class numbers,** *Proc. Amer. Math. Soc.* **31,** No. 1, 51-53 (Jan. 1972).

12705. Reichard, T. W., **Mechanical properties of insulating concretes,** Paper in *American Concrete Spec. Publ. SP 29, Lightweight Concrete,* pp. 253-316 (American Concrete Institute, Detroit, Mich., 1971).

12706. Reimann, C. W., **The single crystal spectra of dichlorotetrapyrazolenickel(II), dibromotetrapyrazolenickel(II), and hexapyrazolenickel(II) nitrate,** *J. Chem. Phys.* **73,** No. 3, 561-568 (Feb. 5, 1970).

12707. Ritter, J. J., Coyle, T. D., Bellama, J. M., **Synthesis of ethynylboron halides,** *Chem. Commun.,* pp. 908-909 (1969).

12708. Roberts, J. R., Voigt, P. A., **The continuous emission from hydrogen at the bound-to-free transition regions,** *Proc. Xth Int. Conf. on Phenomena in Ionized Gases, Oxford, Eng., Sept. 13-18, 1971,* p. 375 (1971).

12709. Rosenblatt, J. R., Filliben, J. J., **Randomization and the draft lottery,** *Science* **171,** 306-308 (Jan. 22, 1971).

12710. Rossmassler, S. A., **Federal information programs,** *Chem. Er.g. Prog.* **67,** No. 11, 75-76 (Nov. 1971).

12711. Sengers, J. M. H. Levelt, Chen, W. T., **Vapor pressure, critical isochore, and some metastable states of CO_2,** *J. Chem. Phys.* **56,** No. 1, 595-608 (Jan. 1, 1972).

12712. Simmons, J. D., McDonald, J. K., **The emission spectrum of AlN,** *J. Mol. Spectrosc.* **41,** 584-594 (1972).

12713. Taylor, J. K., Ed., Alvarez, R., Paulson, R. A., Rains, T. C., Rook, H. L., **Interaction of nitrilotriacetic acid with suspended and bottom material,** *Water Pollution Control Research Series 16020 GFR 07-71—Environmental Protection Agency (Progress Report),* pp. 1-31 (July 1971).

12714. Thom, H. C. S., Marshall, R. D., **Wind and surge damage due to hurricane Camille,** *J. Waterways, Harbors, Coastal Eng. Div. Proc. Amer. Soc. Civil Eng.* **WW2,** No. 8144, 355-363 (May 1971).

12715. von Busch, F., Dunn, G. H., **Photodissociation of H_2^+ and D_2^+: Experiment,** *Phys. Rev. A* **5,** No. 4, 1726-1743 (Apr. 1972).

12716. Scheer, M. D., Klein, R., **The addition of O(^{3}P) to olefins. The nature of the intermediate,** *J. Phys. Chem.* **74,** 2732-2733 (1970).

12717. Yakowitz, H., Ballantyne, J. P., Munro, E., Nixon, W. C., **The cylindrical secondary electron detector as a voltage measuring device in the scanning electron microscope,** *Proc. 5th Annual Scanning Electron Microscopy Symp., ITT Research Institute, Chicago, Ill., Apr. 1972,* Part 1, 33-40 (1972).

12718. Ballard, D. B., **Comparison and evaluation of specimens for resolution standard,** *Proc. 5th Annual Scanning Electron Microscopy Symp., ITT Research Institute, Chicago, Ill., Apr. 1972,* Part 1, 121-128 (1972).

12719. Rosenstein, M., McLaughlin, W. L., Silverman, J., **Energy deposition of electron beams in insulating materials** (Proc. 4th Int. Conf. on Electron and Ion Beam Science and Technology, Los Angeles, Calif., May 1970), Chapter in *Electron and Ion Beam Science and Technology, 4th International Conference*, R. Bakish, ed., pp. 591-604 (The Electrochemical Society, New York, N.Y., 1970).

12720. Powell, C. J., **Structure on the high-energy side of the $KL_{23}M$ Auger peak from solid aluminum: internal photoemission,** *Appl. Phys. Lett.* **20,** No. 9, 335-337 (May 1, 1972).

12721. Carter, G. C., Weisman, I. D., Bennett, L. H., Watson, R. E., **The $AuAl_2$-$AuGa_2$-$AuIn_2$ problem: Knight shifts and relaxation times in their pseudobinary alloys,** *Phys. Rev. B* **5,** No. 9, 3621-3638 (May 1, 1972).

12722. Shapiro, H. M., Bryan, S. D., Lipkin, L. E., Stein, P. G., Lemkin, P. F., **Computer-aided microspectrophotometry of biological specimens,** *Exp. Cell Res.* **67,** No. 1, 81-89 (July 1971).

12723. Semmelroth, C. C., **Adjustment of the Munsell-value and W^*-scales to uniform lightness steps for various background reflectances,** *Appl. Opt.* **10,** No. 1, 14-18 (Jan. 1971).

12724. Bender, P. L., Dicke, R. H., Wilkinson, D. T., Alley, C. O., Currie, D. G., Faller, J. E., Mulholland, J. D., Silverberg, E. C., Plotkin, H. H., Kaula, W. M., MacDonald, G. J. F., **The lunar laser ranging experiment** (Proc. Conf. on Experimental Tests of Gravitational Theories, Pasadena, Calif., Nov. 1970), *JPL Technical Memo* **33-499,** pp. 178-181 (Jet Propulsion Laboratories, Pasadena, Calif., Nov. 1971).

12725. Semmelroth, C. C., **Prediction of lightness and brightness on different backgrounds,** *J. Opt. Soc. Amer.* **60,** No. 12, 1685-1689 (Dec. 1970).

12726. Mahler, R. J., Phelan, R. J., Jr., Cook, A. R., **High D^*, fast, lead zirconate titanate pyroelectric detectors,** *Infrared Phys.* **12,** 57-59 (1972).

12727. Mulholland, J. D., Alley, C. O., Bender, P. L., Currie, D. G., Dicke, R. H., Faller, J. E., Kaula, W. M., MacDonald, G. J. F., Plotkin, H. H., Wilkinson, D. T., **Preliminary results of laser ranging to a reflector on the lunar surface** (Proc. 13th Plenary Meeting of COSPAR, Leningrad, May 1970), *Space Research XI*, 97-104 (Academy-Verlag, Berlin, 1971).

12728. Sengers, J. V., **Triple collision effects in the transport properties for a gas of hard spheres** (Proc. Symp. on Kinetic Equations, Cornell University, Ithaca, N.Y., 1969), Paper C-1 in *Kinetic Equations*, R. L. Kiboff and N. Rostoker, eds., pp. 137-193 (Gordon and Breach, New York, N.Y., 1971).

12729. Danos, M., **Baryon resonances in the nuclear ground state: How about double counting?** (Proc. 3d Int. Conf. High-Energy Physics and Nuclear Structure, Columbia University, New York, N.Y., Sept. 7-12, 1969), Chapter in *High-Energy Physics and Nuclear Structure*, pp. 811-815 (Plenum Press, New York, N.Y., 1970).

12730. Flynn, D. R., **Considerations in launching effective noise control programs** (Proc. EASCON 1971 Conf., Oct. 6-8, 1971, Washington, D.C.), *IEEE Trans. Aerosp. Electron. Syst.*, pp. 53-59 (Nov. 1971).

12731. Nimeroff, I., **The psychology of color,** *Amer. Paint J.*, pp. 65-70 (Jan. 11, 1971).

12732. Yule, H. P., **Study of gamma-ray spectrum distortion by mathematical smoothing,** *Anal. Chem.* **44,** No. 7, 1245-1249 (June 1972).

12733. Rook, H. L., Gills, T. E., LaFleur, P. D., **Method for determination of mercury in biological materials by neutron activation analysis,** *Anal. Chem.* **44,** No. 7, 1114-1117 (June 1972).

12734. Rook, H. L., **Rapid, quantitative separation for the determination of selenium using neutron activation analysis,** *Anal. Chem.* **44,** No. 7, 1276-1278 (June 1972).

12735. Yates, J. T., Jr., Madey, T. E., **The adsorption of methane by tungsten (100),** *Surface Sci.* **28,** 437-459 (1971).

12736. Madey, T. E., **Adsorption and displacement processes on W(111) involving CH_4, H_2, and O_2,** *Surface Sci.* **29,** 571-589 (1972).

12737. Comeford, J. J., **The spectral distribution of radiant energy of a gas-fired radiant panel and some diffusion flames,** *Combust. Flame* **18,** 125-132 (1972).

12738. LaVilla, R. E., **$K\alpha$ emission spectrum of gaseous N_2,** *J. Chem. Phys.* **56,** No. 5, 2345-2349 (Mar. 1, 1972).

12739. Okabe, H., **Photodissociation of $OSCl_2$ in the vacuum ultraviolet; the heat of formation of SO,** *J. Chem. Phys.* **56,** No. 7, 3378-3381 (Apr. 1, 1972).

12740. Stephenson, N. C., Roth, R. S., **The crystal structure of the high temperature form of Ta_2O_5,** *J. Solid State Chem.* **3,** 145-153 (1971).

12741. Schrack, R. A., Heaton, H. T., II, Schwartz, R. B., **Suppression of after-pulsing in the type 58AVP photomultiplier,** *Nucl. Instr. Methods* **77,** 175-176 (1970).

12742. Thompson, B. A., Lutz, G. J., **Characterization of ancient bronze artifacts by neutron and photon activation analysis,** *Radiochem. Radioanal. Lett.* **9,** No. 5-6, 343-350 (Mar. 1972).

12743. Seltzer, S. M., Berger, M. J., **Response of NaI detectors to high-energy gamma rays,** *Trans. Amer. Nucl. Soc.* **14,** No. 1, 124-125 (June 1971).

12744. Sanchez, I. C., DiMarzio, E. A., **Dilute-solution theory of polymer crystal growth. Some thermodynamic and predictive aspects for polyethylene,** *Macromolecules* **4,** 677-687 (Nov.-Dec. 1971).

12745. Garvin, D., Evans, W. H., Duncan, B. C., **Some information processing techniques for the small, semi-automated, scientifically-oriented data center** (Proc. Forum on the Management of Information Analysis Centers, National Bureau of Standards, Gaithersburg, Md., May 17-19, 1971), Chapter in *The Management of Information Analysis Centers*, W. A. Smith, ed., **72-1**, 40-57 (Sponsored by The COSATI Panel Information Analysis Centers, U.S. Atomic Energy Commission, Oak Ridge, Tenn., Jan. 1972).

12746. Bridges, J. M., Richter, J., **Arc measurements of Fe I oscillator strengths with improved accuracy,** *Proc. Xth Int. Conf. on Phenomena in Ionized Gases, Oxford, England, 1971*, p. 384 (Donald Parsons and Company, Ltd., Oxford, England, 1971).

12747. Scheer, M. D., Fine, J., **The positive and negative self-surface ionization of tantalum,** Chapter 39 in *The Structure and Chemistry of Solid Surfaces*, G. A. Somorjai, ed., pp. 39-1—39-15 (John Wiley & Sons, Inc., New York, N.Y., 1969).

12748. Milligan, D. E., Jacox, M. E., **Infrared and ultraviolet spectroscopic studies of free radicals and molecular ions isolated in inert solid matrices,** Paper in *Molecular Spectroscopy: Modern Research*, pp. 259-286 (Academic Press, Inc., New York, N.Y., 1972).

12749. Moore-Sitterly, Ch., **Committee 4: Structure of atomic spectra** (Commission 14, International Astronomical Union Report), *Trans. I.A.U.* **14A,** 133-137 (1970).

12750. Swanson, N., Kuyatt, C. E., Cooper, J. W., Krauss, M., **Alternative decay channels of a CO^- Feshbach resonance,** *Phys. Rev. Lett.* **28,** No. 15, 948-951 (Apr. 10, 1972).

12751. Pope, C. I., **A simplified method for determining residual thiosulfate in processed microfilm,** *Photogr. Sci. Eng.* **13,** No. 5, 278-279 (Sept.-Oct. 1969).

12752. Schubauer, G. B., **Early developments in hot-wire anemometry at NBS and a look at some elsewhere** (Proc. Int. Symp. on Hot-Wire Anemometry, University of Maryland, College Park, Md., Mar. 20-21, 1967), Chapter in *Advances in Hot-Wire Anemometry*, W. L. Melnik and J. R. Weske, eds., pp. 13-24 (July 1968).

12753. Smith, M. W., Wiese, W. L., **Graphical presentations of systematic trends of atomic oscillator strengths along isoelectronic sequences and new oscillator strengths derived by interpolation,** *Astrophys. J. Supplement Series* **23,** Supplement No. 196, 103-196 (June 1971).

12754. Simmons, J. H., Macedo, P. B., **Shear and volume structural relaxation in immiscible oxide melts** (Proc. Conf. Physics of Non-Crystalline Solids, Sheffield, England, Sept. 1970), Chapter 7 in *Amorphous Materials*, R. W. Douglas and B. Ellis, eds., pp. 69-78 (John Wiley & Sons, London, England, 1972).

12755. Rosasco, G. J., Benoit, C., Weber, A., **Light scattering study of the attenuation and dispersion of hypersound in ammonium bromide** (Proc. 2d Int. Conf. on Light Scattering in Solids, Paris, France, July 19-23, 1971), Chapter in *Light Scattering in Solids*, M. Balkanski, ed., pp. 483-488 (Flammarion Sciences Publ., Paris, France, 1971).

12756. Berger, M. J., Seltzer, S. M., Eisen, H. A., Silverman, J., **Absorption of electron energy in multi-layer targets,** *Trans. Amer. Nucl. Soc.* **14,** No. 1, 887-888 (June 1971).

12757. Sher, A. H., Keery, W. J., Dyson, H. E., **Improved infrared response measurements in semiconductor nuclear radiation detectors,** *IEEE Trans. Nucl. Sci.* **NS-19,** No. 1, 341-344 (Feb. 1972).

12758. Sher, A. H., Keery, W. J., **Improved infrared-response technique for determining impurity and defect levels in semiconductors,** *Appl. Phys. Lett.* **20,** No. 3, 120-122 (Feb. 1, 1972).

12759. Stephenson, N. C., Roth, R. S., **Structural systematics in the binary system Ta_2O_5 - WO_3. V. The structure of the low-temperature form of tantalum oxide L-Ta_2O_5,** *Acta Cryst.* **B27,** 1037-1044 (1971).

12760. Stephenson, N. C., Roth, R. S., **Structural systematics in the binary system Ta_2O_5-WO_3. IV. The structure of $Ta_{38}WO_{94}$,** *Acta Cryst.* **B27,** 1031-1036 (1971).

12761. Stephenson, N. C., Roth, R. S., **Structural systematics in the binary system Ta_2O_5-WO_3. III. The structure of $45Ta_2O_5 \cdot Al_2O_3 \cdot 4WO_3$,** *Acta Cryst.* **B27,** 1025-1031 (1971).

12762. Stephenson, N. C., Roth, R. S., **Structural systematics in the binary system Ta_2O_5-WO_3. II. The structure of $Ta_{30}W_2O_{81}$,** *Acta Cryst.* **B27,** 1018-1024 (1971).

12763. Stephenson, N. C., Roth, R. S., **Structural systematics in the binary system Ta_2O_5-WO_3. I. The structure of $Ta_{22}W_4O_{67}$,** *Acta Cryst.* **B27,** 1010-1017 (1971).

12764. Eisenhauer, C., Chilton, A. B., **Angular distribution of scattered gamma rays from a fan source,** *Trans. Amer. Nucl. Soc.* **14,** No. 1, 402-403 (June 1971).

12765. Sher, A. H., Liu, Y. M., Keery, W. J., **Infrared response measurements on radiation-damaged Si(Li) detectors,** *IEEE Trans. Nucl. Sci.* **NS-19,** No. 3, 312-317 (1972).

12766. Meinke, W. W., **Some comments on activation analysis,** *Proc. 2d Inter-American Conference on Radiochemistry, Mexico City, Mexico, Apr. 22-25, 1968*, pp. 54-60 (Published by Secretaría General de la Organización de Estados Americanos, Washington, D.C., 1971).

12767. Smith, C. N., **Limit of error calculations used in integrated safeguards experiment,** *Proc. 12th Annual Meeting of the Institute of Nuclear Materials Management, Palm Beach Shores, Fla., June 28-July 1, 1971,* **1,** 193-207 (June 1971).

12768. Geist, J., Kendall, J. M., Sr., **Circumsolar radiation and the international pyrheliometric scale,** *Appl. Opt.* **11,** No. 6, 1437-1439 (June 1972).

12769. Rasberry, S. D., **Application of computers in electron probe and x-ray fluorescence analysis** (Proc. 20th Annual Conf. on Applications of X-Ray Analysis, Denver, Colo., Aug. 11-13, 1971), Chapter in *Advances in X-Ray Analysis*, K. F. J. Heinrich, ed., **15,** 56-69 (Plenum Press, Inc., New York, N.Y., 1972).

12770. Sugar, J., Kaufman, V., **Fourth spectrum of lutetium,** *J. Opt. Soc. Amer.* **62,** No. 4, 562-570 (Apr. 1972).

12771. Schaffer, R., **Documentation interest of the National Bureau of Standards and collaborating groups for communication of laboratory results in clinical analysis** (Proc. Conf. Clinically Oriented Documentation of Laboratory Data, Buffalo, N.Y., May 10-12, 1971), Chapter in *Clinically Oriented Documentation of Laboratory Data*, E. R. Gabrieli, ed., pp. 419-424 (Academic Press, Inc., New York and London, 1972).

12772. Schaffer, R., **Isotopic methods (of carbohydrate analysis),** Paper in *The Carbohydrates, Chemistry and Biochemistry, Second Edition*, W. Pigman and D. Horton, eds., **11B,** 765-776 (Academic Press, Inc., New York and London, 1970).

12773. Shafer, M. R., **Discussion of ASME paper 64-WA/FM-1, density effect and Reynolds number effect on gas turbine flowmeters by W. F. Z. Lee and H. J. Evans,** *J. Basic Eng. Trans. ASME, Series D.* **87,** No. 4, 1052-1053 (Dec. 1965).

12774. Forman, R. A., Piermarini, G. J., Barnett, J. D., Block, S., **Pressure measurement made by the utilization of ruby sharp-line luminescence,** *Science* **176,** 284-285 (Apr. 21, 1972).

12775. Linsky, J. L., **A proof of the relation between reflectivity and emissivity in an isothermal scattering and absorptive atmosphere,** *J. Quant. Spectrosc. Radiat. Trans.* **12,** No. 5, 777-781 (May 1972).

12776. Roberts, J. R., Voigt, P. A., Nagy, J. J., **Continuous emission from hydrogen plasmas. I. The theoretical hydrogen emission coefficients,** *Proc. 9th Int. Conf. Phenomena in Ionized Gases, Bucharest, Romania, Sept. 1-6, 1969*, p. 595 (Editura Academiei Republicii Socialiste Romania, Bucharest, Romania, 1969).

12777. Coxon, B., **Conformational analysis via nuclear magnetic resonance spectroscopy,** Article 93 in *Methods in Carbohydrate Chemistry*, R. L. Whistler and J. N. BeMiller, eds., **VI,** 513-539 (Academic Press, Inc., New York, N.Y., 1972).

12778. Perls, T. A., Stern, J., **Motion,** Chapter 3 in *ISA Transducer Compendium, Second Edition*, Part II, 105-111 (IFI/Plenum, New York, N.Y., June 1970).

12779. Yokel, F. Y., Somes, N. F., **Performance criteria for tall buildings** (Proc. 5th CIB Congress, Paris-Versailles, France, June 1971), Chapter in *Research Into Practice, The Challenge of Application* **1**, 587-592 (June 1971).

12780. Stiehler, R. D., **Letter to the Editor,** *Modern Plastics* **49**, No. 5, 8 (May 1972).

12781. Smith, C. N., **X-ray standards for law enforcement,** Article in *The Police Chief,* p. 46 (International Association of Chiefs of Police, Gaithersburg, Md., May 1972).

12782. Liu, Y. M., Coleman, J. A., **Radiation damage effects by electrons, protons, and neutrons in Si(Li) detectors,** *IEEE Trans. Nucl. Sci.* **NS-19**, No. 3, 346-352 (1972).

17283. Radebaugh, R., Siegwarth, J. D., **Numerical analysis of continuous and discrete heat exchangers for dilution refrigerators** (Proc. Comm. I, Conf. on Cryophysics and Cryoengineering, Tokyo, Japan, Sept. 11-12, 1970), *Bull. Inst. Int. Froid, Annexe 1970-2,* pp. 57-62 (Int. Inst. of Refrigeration, Paris, France, 1970).

12784. Clark, A. F., Kropschot, R. H., **Low temperature specific heat and thermal expansion of alloys** (Proc. Comm. I, Conf. on Cryophysics and Cryoengineering, Tokyo, Japan, Sept. 11-12, 1970), *Bull. Inst. Int. Froid, Annexe 1970-2,* pp. 249-254 (Int. Inst. of Refrigeration, Paris, France, 1970).

12785. Andrews, J. R., **A frequency calibrator for UHF using an avalanche transistor,** *QST* **LVI**, No. 5, 16-18 (May 1972).

12786. Oser, H. J., Translator and editor of the book: *Boundary layers of flow and temperature,* **by A. Walz, 297 pages (MIT Press, Cambridge, Mass., 1969).**

12787. Milligan, D. E., Jacox, M. E., **Infrared and ultraviolet spectroscopic studies of molecular ions trapped in inert solid matrices,** *Proc. 19th Annual Conf. on Mass Spectrometry and Allied Topics, Atlanta, Ga., May 2-7, 1971,* pp. 100-104 (American Society for Mass Spectrometry, ASTM Committee E-14 and Naval Research Laboratory, Washington, D.C., 1971).

12788. Horton, W. S., Carson, C. C., **Gas analysis: Determination of gases in metals,** Chapter 103 in *Treatise on Analytical Chemistry,* I. M. Kolthoff and P. J. Elving, eds., **10,** Part 1, Section E, 6017-6144 (John Wiley & Sons, Inc., New York, N.Y., 1972).

12789. Meyerson, M. R., **Ultrahigh-strength materials,** Chapter 19 in *Materials Science and Technology for Design Engineers,* pp. 366-389 (Hayden Book Company, Inc., New York, N.Y., 1972).

12790. Hougen, J. T., **An interesting example of Hund's coupling case (e) in $HgAr^+$,** *J. Mol. Spectrosc.* **42**, 381-384 (1972).

12791. King, D. A., Madey, T. E., Yates, J. T., Jr., **Interaction of oxygen with polycrystalline tungsten. Part 3—Electron stimulated desorption,** *J. Chem. Soc., Faraday Trans. I,* **68**, 1347-1359 (1972).

12792. Cox, D. M., Smith, S. J., **Absolute measurement of 2*S* atom production by dissociative excitation of molecular deuterium,** *Phys. Rev. A* **5**, No. 6, 2428-2442 (June 1972).

12793. Opal, C. B., Beaty, E. C., **Measurements of large angle inelastic scattering cross sections for electrons on helium,** *J. Phys. B: At. Mol. Phys.* **5**, No. 2, 627-635 (Mar. 1972).

12794. Hidalgo, M. B., Geltman, S., **A high energy approximation: III. Helium excitation by electrons,** *J. Phys. B: At. Mol. Phys.* **5**, No. 3, 617-626 (Mar. 1972).

12795. Albers, J., Mountain, R. D., **Asymptotic form of the pair-correlation function in liquids,** *Phys. Rev. A* **5**, No. 6, 2629-2632 (June 1972).

12796. Maienthal, E. J., **Polarographic analysis at NBS,** *Amer. Lab.* **4**, No. 6, 12-21 (June 1972).

12797. Bay, Z., White, J. A., **Frequency dependence of the speed of light in space,** *Phys. Rev. D* **5**, No. 4, 796-799 (Feb. 15, 1972).

12798. Rathbone, D. Z., Jr., **Cutting the costs of cooling,** *Plant Eng.* **26**, No. 12, 88-90 (June 15, 1972).

12799. Pilling, M. J., Bass, A. M., Braun, W., **A curve of growth determination of the *f*-values for the fourth positive system of CO and the Lyman-Birge-Hopfield system of N_2,** *J. Quant. Spectrosc. Radiat. Trans.* **11**, 1593-1604 (1971).

12800. Hubbell, J. H., **Survey of photon-attenuation-coefficient measurements 10 eV to 100 GeV,** *Atomic Data* **3**, No. 3, 241-297 (Nov. 1971).

12801. Spiegel, V., Jr., **Evaluation of the neutron source absorption correction in a manganous sulfate bath** (Proc. Amer. Nucl. Soc. National Topical Meeting on Neutron Sources and Applications, Augusta, Ga., Apr. 18-21, 1971), Chapter in *Neutron Sources and Applications,* **II**, 1-170—1-171 (1971).

12802. Smith, J. M., **Automatic data evaluation, manipulation, display, and plotting with SPEED,** *Comput. Graphics* **4**, No. 2, 41-53 (Fall 1970).

12803. Karp, S. S., Rath, G. J., Wright, G. P., **A simulation model of admission and discharge policies for an intensive care unit,** *Proc. 1970 Summer Computer Simulation Conference, Denver, Colo., June 10-12, 1970,* pp. 1108-1114 (June 1970).

12804. Slattery, W. J., **Standards: Information and sources,** *Proc. 2d Annual Plant Engineering and Western Maintenance Conf., Anaheim, Calif., Sept. 23-25, 1969,* pp. 49-57 (American Institute of Plant Engineers, Region 6, Cincinnati, Ohio, 1969).

12805. Paffenbarger, G. C., **Dental cements, direct filling resins, composite and adhesive restorative materials: A resume** (Proc. Engineering Foundation Research Conferences, Engineering in Medicine-Bioceramics, New England College, Henniker, New Hampshire, August 3-7, 1970), *J. Biomed. Mater. Res. Symp.* No. 2, Part 2, 363-393 (John Wiley & Sons, Inc., New York, N.Y., Jan. 1972).

12806. Guevara, F. A., McInteer, B. B., Ottesen, D., Hanley, H. J. M., **A critique of the high-temperature viscosity measurements of Trautz and Zink,** Los Alamos Scientific Laboratory, University of California, Los Alamos, New Mexico, Report No. **LA-4643-MS,** pp. 1-10 (June 1971).

12807. Klebanoff, P. S., Tidstrom, K. D., **Mechanism by which a two-dimensional roughness element induces boundary-layer transition,** *Phys. Fluids* **15**, No. 7, 1173-1188 (July 1972).

12808. Greer, W. L., Rubin, R. J., **Quantum theory of heat transport in an isotopically substituted, one-dimensional, harmonic crystal,** *J. Math. Phys.* **13**, No. 3, 379-385 (Mar. 1972).

12809. Jacox, M. E., Milligan, D. E., **Spectrum and structure of the HO_2 free radical,** *J. Mol. Spectry.* **42**, No. 3, 495-513 (June 1972).

12810. Levine, J., Hall, J. L., **Design and operaton of a methane absorption stabilized laser strainmeter,** *J. Geophys. Res.* **77**, No. 14, 2595-2609 (May 10, 1972).

12811. Winogradoff, N. M., Haller, W. K., Hockey, B. J., **Cross-sectional energy distributions in gigawatt laser pulses: Impli-**

cations and determination, *Opto-Electron. Lett.* **3,** 145-148 (1971).

12812. Spencer, L. V., Woolf, S., **A "forced reflection"-type of collimator design,** Paper in Chapter on Shielding and Dosimetry, *Trans. Amer. Nucl. Soc.* **14,** No. 1, 887 (June 1971).

12813. Jednačak, J., Hudomalj, M., Pravdić, V., Haller, W., **Electrokinetics of glass surfaces. I. Charge densities at the solid/gas and the solid/liquid interfaces of porous glasses,** *Croat. Chem. Acta* **44,** 149-156 (1972).

12814. Kuyatt, C. E., Plummer, E. W., **Field emission deflection energy analyzer,** *Rev. Sci. Instr.* **43,** No. 1, 109-111 (Jan. 1972).

12815. Liebman, J. F., **Comment on the paper "Oxygen monofluoride (OF, $^2\Pi$): Hartree-Fock wavefunction, binding energy, ionization potential, electron affinity, dipole and quadrupole moments, and spectroscopic constants. A comparison of theoretical and experimental results,"** by P. A. G. O'Hare and A. C. Wahl, *J. Chem. Phys.* **56,** No. 8, 4242-4243 (Apr. 15, 1972).

12816. Brinckman, F. E., Johannesen, R. B., Handy, L. B., **Pi-bonding effects in organoxytungsten(VI) fluorides,** *J. Fluorine Chem. Short Commun.* **1,** 493-497 (1971/72).

12817. Becker, D. A., LaFleur, P. D., **Determination of trace quantities of uranium in biological materials by neutron activation analysis using a rapid radiochemical separation,** *Anal. Chem.* **44,** No. 8, 1508-1511 (July 1972).

12818. Leung, P. S., Rush, J. J., Taylor, T. I., **Study of hindered rotation in crystals by neutron transmission measurements: Comparison of calculated and measured scattering cross sections,** *J. Chem. Phys.* **57,** No. 1, 175-182 (July 1, 1972).

12819. Prince, E., **Configuration of the water molecule in cupric fluoride dihydrate at 298 K,** *J. Chem. Phys.* **56,** No. 9, 4352-4355 (May 1, 1972).

12820. Kurylo, M. J., **Absolute rate constants for the addition of O(^{3}P) atoms to propylene,** *Chem. Phys. Lett.* **14,** No. 1, 117-120 (May 1, 1972).

12821. Marchetti, M. A., Krauss, M., Weiss, A. W., **Symmetry-adapted pair correlations in O and O^-,** *Phys. Rev. A* **5,** No. 6, 2387-2390 (June 1972).

12822. de Graaf, L. A., Rush, J. J., Flotow, H. E., Rowe, J. M., **Quasielastic thermal neutron scattering by hydrogen in α-vanadium hydride,** *J. Chem. Phys.* **56,** No. 9, 4574-4581 (May 1, 1972).

12823. Davis, D. D., Huie, R. E., Herron, J. T., Kurylo, M. J., Braun, W., **Absolute rate constants for the reaction of atomic oxygen with ethylene over the temperature range 232-500 K,** *J. Chem. Phys.* **56,** No. 10, 4868-4876 (May 15, 1972).

12824. Eby, R. K., Colson, J. P., **Alpha relaxation in polyethylene,** *J. Appl. Phys.* **41,** No. 11, 4313-4316 (Oct. 1970).

12825. Scheer, M. D., Klein, R., McKinley, J. D., **Halogens adsorbed on molybdenum: Their surface lifetimes and desorption kinetics,** *Surface Sci.* **30,** No. 2, 251-262 (Apr. 1972).

12826. Reader, J., Epstein, G. L., **Analysis of the spectrum of quadruply ionized yttrium (Y v),** *J. Opt. Soc. Amer.* **62,** No. 5, 619-622 (May 1972).

12827. Barnes, J. D., Fanconi, B. M., **Raman spectroscopy, rotational isomerism, and the "rotator" phase transition in *n*-alkanes,** *J. Chem. Phys.* **56,** No. 10, 5190-5192 (May 15, 1972).

12828. Okabe, H., **Photodissociation of CS_2 in the vacuum ultraviolet; determination of $D_0°$(SC-S),** *J. Chem. Phys.* **56,** No. 9, 4381-4384 (May 1, 1972).

12829. Camus, P., Sugar, J., **Theoretical study and interpretation of the electronic configurations in Tm II. II. Odd excited configurations $4f^{12}(5d+6s)6p$ and $4f^{13}7s$,** *Phys. Scr.* **4,** 263-268 (1971).

12830. Camus, P., Sugar, J., **Theoretical study and interpretation of the electronic configurations in Tm II. I. Even configurations $4f^{12}(5d+6s)^2+4f^{13}6p$,** *Phys. Scr.* **4,** 257-261 (1971).

12831. Tsang, W., **Comparative rate single-pulse shock tube studies on the thermal decomposition of cyclohexene, 2,2,3-trimethylbutane, isopropyl bromide, and ethylcyclobutane,** *Int. J. Chem. Kinet.* **II,** 311-323 (1970).

12832. Tsang, W., **Rate and mechanism of thermal decomposition of 4-methyl-1-pentyne in a single-pulse shock tube,** *Int. J. Chem. Kinet.* **II,** 23-36 (1970).

12833. Ehrlich, M., **Response of thermoluminescent dosimeters to 15 MeV electrons and ^{60}Co gamma rays,** *Health Phys.* **18,** 287-289 (1970).

12834. Wilson, W. K., Hebert, R. L., **Evaluation of the stability of manifold papers,** *Tappi* **55,** No. 7, 1103-1107 (July 1972).

12835. Schaffer, R., **Occurrence, properties, and preparation of naturally occurring monosaccharides (including 6-deoxy sugars),** Chapter in *The Carbohydrates. Chemistry and Biochemistry*, Second Edition, W. Pigman and D. Horton, eds., **1A,** 69-111 (Academic Press, Inc., New York, N.Y., 1972).

12836. Young, R., Ward, J., Scire, F., **Observation of metal-vacuum-metal tunneling, field emission, and the transition region,** *Phys. Rev. Lett.* **27,** No. 14, 922-924 (Oct. 4, 1971).

12837. Young, R. D., **Surface microtopography,** *Phys. Today* **24,** No. 11, 42-49 (Nov. 1971).

12838. Berger, M. J., Seltzer, S. M., **Bremsstrahlung in the atmosphere,** *J. Atmos. Terr. Phys.* **34,** 85-108 (1972).

12839. Sugar, J., **Potential-barrier effects in photoabsorption. II. Interpretation of photoabsorption resonances in lanthanide metals at the 4*d*-electron threshold,** *Phys. Rev. B* **5,** No. 5, 1785-1792 (Mar. 1, 1972).

12840. Srinivasan, G. R., Tweer, I., Macedo, P. B., Sarkar, A., Haller, W., **Phase separation in SiO_2-B_2O_3-Na_2O system,** *J. Non-Cryst. Solids* **6,** 221-239 (1971).

12841. Fuller, E. G., **Nuclear reaction theory and the photonuclear giant resonance** (Proc. 5th Congress of Yugoslavian Mathematicians, Physicists and Astronomers, Skopje, Yugoslavia, Sept. 13-14, 1972), *Physics* **II,** No. 5, 5-10 (Macedonian Society of Mathematicians, Physicists and Astronomers, Skopje, Yugoslavia, 1972).

12842. Goldberg, R. N., Frank, H. S., **Liquid junction potentials and single-ion activities by computer simulation. I. The concentration cell with transference,** *J. Phys. Chem.* **76,** No. 12, 1758-1762 (1972).

12843. Currie, L. A., **On the interpretation of errors in counting experiments,** *Anal. Lett.* **4,** No. 12, 873-882 (1971).

12844. Carpenter, B. S., **Boron and uranium determination in simulated lunar glass by the nuclear track technique** (Proc. American Nuclear Society 16th Annual Meeting, Los Angeles, Calif., June 28-July 2, 1970), *Trans. Amer. Nucl. Soc.* **13,** No. 1, 79 (1970).

12845. Kirstein, A. F., **Discussion of the paper "Behavior of fiber glass reinforced concrete beams,"** E. G. Nawy, G. E. Neuwerth, and C. J. Phillips, eds., *J. Struct. Div. Proc. Amer. Soc. Civil Eng.* **98,** No. ST5, 1206-1207 (May 1972).

12846. Plummer, E. W., Bell, A. E., **Field emission energy distributions of hydrogen and deuterium on the (100) and (110) planes of tungsten,** *J. Vac. Sci. Technol.* **9,** No. 2, 583-590 (Mar.-Apr. 1972).

12847. Gadzuk, J. W., **Tunneling from Cambridge surface states,** *J. Vac. Sci. Technol.* **9,** No. 2, 591-596 (Mar.-Apr. 1972).

12848. Marzetta, L. A., **An evaluation of the three-voltmeter method for ac power measurement** (Proc. Conf. Digest on Precision Electromagnetic Measurements, Boulder, Colo., June 26, 1972), *IEEE Trans. Instr. Meas.*, pp. 41-42 (June 26, 1972).

12849. Lehman, D. R., **Quasielastic electron scattering from ^{3}He and ^{3}H,** *Phys. Rev. C* **3,** No. 5, 1827-1840 (May 1971).

12850. Menke, J. L., **A fast pulse-dc integrator,** *Nucl. Instrum. Methods* **64,** 61-68 (1968).

12851. Shah, J. A., Danos, M., **Study of the neutron-proton correlations in the shell model,** *Phys. Rev.* **183,** No. 4, 899-907 (July 20, 1969).

12852. Casella, R. C., **Time reversal symmetry and its breaking in physics,** *Phys. Teacher* **8,** No. 3, 114-122 (Mar. 1970).

12853. Lightbody, J. W., Jr., Penner, S., **Electroexcitation of the C^{12} giant resonance,** *Phys. Rev. Lett.* **24,** No. 6, 274-276 (Feb. 9, 1970).

12854. Mohan, R., Danos, M., Biedenharn, L. C., **Three-fluid hydrodynamical model and the proton distribution in ^{208}Pb,** *Phys. Rev. Lett.* **23,** No. 15, 868-870 (Oct. 1969).

12855. Wyckoff, J. M., **Integrated cross sections for ^{209}Bi(γ,*xnyp*) reactions,** *Phys. Rev.* **159,** No. 4, 953-954 (July 20, 1967).

12856. Wyckoff, J. M., **Photon energy emission from spallation products produced by 3-13 GeV electrons,** *Health Phys.* **18,** 693-704 (June 1970).

12857. Praderie, F., Thomas, R. N., **On estimates of the nonradiative energy input to the solar chromosphere from the H^{-} emission,** *Astrophys. J.* **172,** No. 2, 485-490 (Mar. 1, 1972).

12858. O'Toole, W. E., III, Keplinger, M. S., **An experimental program for automatic indexing at the National Bureau of Standards,** *Proc. American Society for Information Science, Philadelphia, Pa., Oct. 11-15, 1970,* **7,** 301-308 (American Society for Information Science, Washington, D.C., Oct. 1970).

12859. Holt, H. K., Sellin, I. A., **Time-dependent theory of Stark-quenching of 2S states in hydrogen and helium,** *Phys. Rev. A* **6,** No. 1, 508-512 (July 1972).

12860. Wyckoff, H. O., Rossi, H. H., **Dosimetric fundamentals and irradiation limits. 1.1. Radiation quantities and units,** Chapter 1 in *Engineering Compendium on Radiation Shielding* **1,** 1-5, (Springer-Verlag, New York, N.Y., 1968).

12861. Wachtman, J. B., Jr., Capps, W., Mandel, J., **Biaxial flexure tests of ceramic substrates,** *J. Mater.* **7,** No. 2, 188-194 (1972).

12862. Hanley, H. J. M., Klein, M., **Application of the *m*-6-8 potential to simple gases,** *J. Phys. Chem.* **76,** No. 12, 1743-1751 (1972).

12863. Spencer, L. V., **On the accuracy of transport calculations using moments methods,** *Trans. Amer. Nucl. Soc.* **14,** No. 1, 339 (June 1971).

12864. Leiss, J. E., **Beam loading and transient behavior in traveling wave electron linear accelerators,** Chapter in *Linear Accelerators,* P. M. Lapostolle and A. L. Septier, eds., pp. 147-172 (North-Holland Publ. Co., Amsterdam, The Netherlands, 1970).

12865. Swartzendruber, L. J., Evans, B. J., **Hyperfine fields at the antimony sites in the Heusler-type Ni_2MnSb alloys** (Proc. 17th AIP Conf. on Magnetism and Magnetic Materials, Chicago, Ill., Nov. 15-18, 1971), Chapter in *Magnetism and Magnetic Materials,* C. D. Graham, Jr., and J. J. Rhyne, eds., No. 5, 539-543 (June 1972).

12866. Swartzendruber, L. J., Bennett, L. H., Kinsman, K. R., **Crystallographic and magnetic transformations of Fe precipitates in Cu** (Proc. 17th AIP Conf. on Magnetism and Magnetic Materials, Chicago, Ill., Nov. 15-18, 1971), Chapter in *Magnetism and Magnetic Materials,* C. D. Graham, Jr., and J. J. Rhyne, eds., No. 5, 408-412 (June 1972).

12867. Sharp, K. G., Coyle, T. D., **Perfluoro(alkylsilanes). II. Trifluoro(trifluoromethyl)silane and trifluoro(pentafluoroethyl)silane,** *Inorg. Chem.* **11,** 1259-1264 (1972).

12868. McMaster, W. H., Del Grande, N. K., Mallett, J. H., Hubbell, J. H., **Compilation of x-ray cross sections,** *Sec. II, Rev. 1, University of California Radiation Laboratory Report 50174, 353 pages,* Lawrence Livermore Laboratory, Livermore, Calif., (May 1969).

12869. Dikkers, R. D., Yokel, F. Y., **Strength of brick walls subject to axial compression and bending** (Proc. 2d International Brick Masonry Conf., Stoke-on-Trent, England, Apr. 12-15, 1970), Chapter in *SIBMAC Proceedings,* H. W. H. West and K. H. Speed, eds., No. 19, 125-132 (The British Ceramic Research Association, Stoke-on-Trent, England, 1971).

12870. Rowen, J. W., Murphey, W. M., Smith, C. N., **Assessment of material unaccounted for control criteria** (Proc. Symp. Progress in Safeguards Techniques, International Atomic Energy Agency, Vienna, Austria, 1970), Chapter in *Safeguards Techniques* **II,** 291-310 (Dec. 1970).

12871. Kidnay, A. J., Miller, R. C., Hiza, M. J., **Solid-vapor and liquid-vapor phase equilibria for the helium-krypton system,** *Ind. Eng. Chem. Fundam.* **10,** No. 3, 459-465 (Aug. 1971).

12872. Eisenhauer, C., **Some current problems in OCD fallout shielding calculations,** *Proc. Special Sessions on Gamma-Ray Production and Transport and on Civil Defense Shielding, Winter Meeting American Nuclear Society, Chicago, Ill., Nov. 1967, A. E. Straker, ed.,* **ANS-SD-7,** (TID) 4500, pp. 143-156 (1969).

12873. Johnson, W. T. K., Dick, C. E., **Half-life measurement of several short-lived nuclear isomers,** *Nucl. Instrum. Methods* **99,** No. 2, 221-226 (1972).

12874. Kondo, A., Brenner, F. C., **Research for a uniform quality grading system for tires. VI. Comparison of the effect of front and rear wheel drive vehicles on projected tread wear of a tire,** *Rubber Chem. Technol.* **44,** No. 4, 960-961 (Sept. 1971).

12875. Linzer, M. A., **A frequency swept injection locked nuclear induction spectrometer with signal averaging,** *Rev. Sci. Instrum.* **43,** No. 5, 723-726 (May 1972).

12876. Treu, S., **A conceptual framework for the searcher-system interface,** *Proc. Workshop on Interactive Bibliographic Search: The User/Computer Interface, Palo Alto, Calif., Jan. 14-15, 1971,* D. E. Walker, ed., pp. 53-66 (AFIPS Press, Montvale, N.J., 1971).

12877. Petrescu-Prahova, I., Winogradoff, N. N., **Photoluminescence in compensated *n*-type, Si-doped gallium arsenide,** *J. Phys. Chem. Solids* **32,** 1873-1879 (1971).

12878. Skopik, D. M., Dodge, W. R., **^{4}He*(e,d)*e'^{2}H angular-distribution and cross-section measurements between 36 and 50 MeV,** *Phys. Rev. C* **6,** No. 1, 43-53 (July 1972).

12879. Robertson, A. F., **Discussion of paper "Fire and the curtain wall"** by H. L. Malhotra, United Kingdom, (Proc. Symp. of CIB Conf. on Curtain Walls, Paris, France, Mar. 3-6, 1970), *Symp. Papers in Light Facades*, pp. 74-75 (June 1971).

12880. Caswell, R. S., **Review of measurements of absolute neutron emission rates and spectra from neutron sources** (Proc. Amer. Nucl. Soc. National Topical Meeting on Neutron Sources and Applications, Augusta, Ga., Apr. 19-21, 1971), Chapter in *Neutron Sources and Applications, CONF-710402*, **I,** 1-53—1-62 (1971).

12881. Jacox, M. E., Milligan, D. E., **Vibrational and electronic spectra of the O_3^- anion isolated in an argon matrix,** *J. Mol. Spectrosc.* **43,** No. 1, 148-167 (July 1972).

12882. Celotta, R. J., Bennett, R. A., Hall, J. L., Siegel, M. W., Levine, J., **Molecular photodetachment spectrometry. II. The electron affinity of O_2 and the structure of O_2,** *Phys. Rev. A* **6,** No. 2, 631-642 (Aug. 1972).

12883. Siegel, M. W., Celotta, R. J., Hall, J. L., Levine, J., Bennett, R. A., **Molecular photodetachment spectrometry. I. The electron affinity of nitric oxide and the molecular constants of NO^-,** *Phys. Rev. A* **6,** No. 2, 607-631 (Aug. 1972).

12884. Utton, D. B., Tsang, T., **Coupled nuclear spin relaxation and internal rotations in magnesium fluosilicate hexahydrate,** *J. Chem. Phys.* **56,** No. 1, 116-119 (Jan. 1, 1972).

12885. Hutchinson, J. M. R., Garfinkel, S. B., **Standardization of cadmium-109 sources for γ-ray emission rate,** *Int. J. Appl. Radiat. Isotop.* **22,** 405-414 (1971).

12886. Hutchinson, J. M. R., Lantz, J. L., Mann, W. B., Mullen, P. A., Rodríguez-Pasqués, R. H., **An anti-coincidence shielded NaI(Tl) system at NBS,** *IEEE Trans. Nucl. Sci.* **NS-19,** No. 1, 117-118 (Feb. 1972).

12887. Rodríguez-Pasqués, R. H., **Microactivity radiochemistry,** *Proc. 2d Inter-American Conference on Radiochemistry, Mexico City, Mexico, Apr. 22-25, 1968*, pp. 278-289 (Published by Secretaría General de la Organización de Estados Americanos, Washington, D.C., 1971).

12888. Hirshfeld, A. T., Hoppes, D. D., Mann, W. B., Schima, F. J., **Nuclear orientation of ^{82}Br in iron** (Proc. Int. Conf. on Hyperfine Interactions in Excited Nuclei, Rehovot, Israel, Sept. 6-11, 1970), Chapter in *Hyperfine Interactions in Excited Nuclei* **1,** 335-338 (Gordon and Breach Sci. Publ., New York, London, Paris, 1971).

12889. Soulen, R. J., Jr., Colwell, J. H., **The equivalence of the superconducting transition temperature of pure indium as determined by electrical resistance, magnetic susceptibility, and heat-capacity measurements,** *J. Low Temp. Phys.* **5,** No. 3, 325-333 (1971).

12890. Cook, C. C., Allred, C. M., **An excitation system for piston attenuators,** *IEEE Trans. Instrum. Meas.* **IM-20,** No. 1, 10-16 (Feb. 1971).

12891. Gimmestad, G. G., Pardoe, G. W. F., Gebbie, H. A., **The study of dimeric molecules in ammonia vapour of submillimeter wave spectroscopy,** *J. Quant. Spectrosc. Radiat. Transfer* **12,** 559-567 (1972).

12892. Zimmerman, J. E., **Josephson effect devices and low-frequency field sensing,** *Cryogenics* **12,** No. 1, 19-31 (Feb. 1972).

12893. Ruff, A. W., Jr., **Interpretation of x-ray diffraction data with regard to stacking faults in silver-tin alloys,** *J. Appl. Phys. Commun.* **43,** No. 6, 2909-2911 (June 1972).

12894. Siegwarth, J. D., Radebaugh, R., **The design of optimum heat exchangers for dilution refrigerators,** *Rev. Sci. Instrum.* **43,** No. 2, 197-204 (Feb. 1972).

12895. Mizushima, M., Evenson, K. M., Wells, J. S., **Laser magnetic resonance of the NO molecule using 78-, 79-, and 119-μm H_2O laser lines,** *Phys. Rev. A* **5,** No. 5, 2276-2287 (May 1972).

12896. Prydz, R., Goodwin, R. D., **Experimental melting and vapor pressures of methane,** *J. Chem. Thermodyn.* **4,** No. 1, 127-133 (Jan. 1972).

12897. Ledbetter, H. M., **An approximation of $\theta_{elastic}$ for cubic metals,** *J. Phys. F* **2,** No. 2, 262-269 (Mar. 1972).

12898. McDonald, D. G., Risley, A. S., Cupp, J. D., Evenson, K. M., Ashley, J. R., **Four-hundredth-order harmonic mixing of microwave and infrared laser radiation using a Josephson junction and a maser,** *Appl. Phys. Lett.* **20,** No. 8, 296-299 (Apr. 15, 1972).

12899. Ledbetter, H. M., Wayman, C. M., **The crystallography of the β-to-β″ AuCd martensitic transformation,** *Acta Met.* **20,** No. 1, 19-30 (Jan. 1972).

12900. Manning, J. R., **Impurity-correlation factors for diffusion in diamond-structure crystals,** *Phys. Rev. B* **6,** No. 4, 1344-1355 (Aug. 15, 1972).

12901. Foote, W. J., Hunter, R. D., **Improved gearing for rotary-vane attenuators,** *Rev. Sci. Instrum.* **43,** No. 7, 1042-1043 (July 1972).

12902. Swartzendruber, L. J., Bennett, L. H., **Retained austenite developed during surface grinding of a carbon steel,** *Scr. Met.* **6,** 737-742 (Aug. 1972).

12903. Hyman, A., Perloff, A., **The crystal structure of bismuth (2:1) borate, $2Bi_2O_3 \cdot B_2O_3$,** *Acta Crystallogr.* **B28,** Part 7, 2007-2011 (1972).

12904. Perloff, A., **The crystal structures of hydrated calcium and ammonium monofluorophosphates: $CaPO_3F \cdot 2H_2O$ and $(NH_4)_2PO_3F \cdot H_2O$,** *Acta Crystallogr.* **B28,** Part 7, 2183-2191 (1972).

12905. Cezairliyan, A., **Measurement of melting point and electrical resistivity (above 3600 K) of tungsten by a pulse heating method,** *High Temp. Sci.* **4,** No. 3, 248-252 (June 1972).

12906. Rubin, R. J., **Velocity autocorrelation function of a large cluster of particles in a one-dimensional crystal,** *J. Chem. Phys.* **57,** No. 1, 312-316 (July 1, 1972).

12907. Vidal, C. R., Hessel, M. M., **Heat-pipe oven for homogeneous mixtures of saturated and unsaturated vapors; application to NaLi,** *J. Appl. Phys.* **43,** No. 6, 2776-2780 (June 1972).

12908. Paule, R. C., **Low level calibration mixtures for gaseous pollutants,** *Anal. Chem.* **44,** No. 8, 1537-1539 (July 1972).

12909. Rubin, R. J., **Span of a polymer chain,** *J. Chem. Phys.* **56,** No. 12, 5747-5757 (June 15, 1972).

12910. Jones, M. C., Palmer, D. C., Tien, C. L., **Infrared absorptivities of transition metals at room and liquid-helium temperatures,** *J. Opt. Soc. Amer.* **62,** No. 3, 353-360 (Mar. 1972).

12911. Greenspan, M., **Acoustic transmission line: Some impedance properties,** *J. Acoust. Soc. Amer. Letters to Editor* **52,** No. 1, Part 2, 455-457 (July 1972).

12912. Smith, R. V., **Two-phase, two-component critical flow in a venturi,** *J. Basic Eng. Trans. ASME, Series D* **94,** No. 1, 147-155 (Mar. 1972).

12913. Patrician, T. J., Ledbetter, H. M., Reed, R. P., **On the apparent rotation of transformation twins in Fe-Ni martensites,** *Metallurgical Trans.* **3,** No. 4, 947-952 (Apr. 1972).

12914. Dickens, B., Bowen, J. S., Brown, W. E., **A refinement of the crystal structure of $CaHPO_4$ (synthetic monetite),** *Acta Crystallogr.* **B28,** Part 3, 797-806 (Mar. 1972).

12915. Bowen, R. L., Argentar, H., **A stabilizing comonomer: I. Synthesis and confirmation of structure,** *J. Dent. Res.* **51,** No. 4, 1071-1074 (July-Aug. 1972).

12916. Paffenbarger, G. C., **Guide to Dental Materials and Devices: A historical review,** *J. Amer. Dent. Ass.* **84,** 1333-1335 (June 1972).

12917. Mihalas, D., Hummer, D. G., Conti, P. S., **On the N III λλ4640, 4097 lines in *Of* stars,** *Astrophys. J.*, **175,** No. 2, Part 2, L99-L104 (July 15, 1972).

12918. Peterson, R. L., **Pseudoresonances in the magnetophonon structure of nondegenerate semiconductors,** *Phys. Rev. Lett.* **28,** No. 7, 431-434 (Feb. 14, 1972).

12919. Young, R., Ward, J., Scire, F., **The topografiner: An instrument for measuring surface microtopography,** *Rev. Sci. Instrum.* **43,** No. 7, 999-1011 (July 1972).

12920. Opal, C. B., Beaty, E. C., Peterson, W. K., **Tables of secondary-electron-production cross sections,** *At. Data* **4,** No. 3, 209-253 (July 1972).

12921. Dickens, B., Bowen, J. S., **Refinement of the crystal structure of $Ca(H_2PO_4)_2 \cdot H_2O$,** *Acta Crystallogr.* **B27,** Part 11, 2247-2255 (Nov. 1971).

12922. Bowen, R. L., Argentar, H., **Tertiary aromatic amine accelerators with molecular weights above 400,** *J. Dent. Res.* **51,** No. 2, 473-482 (Mar.-Apr. 1972).

12923. Brauer, G. M., Huget, E. F., **Dental adhesives,** Chapter 16 in *The Chemistry of Biosurfaces,* M. L. Hair, ed., **2,** 731-800 (Marcel Dekker, New York, N.Y., 1972).

12924. Jefferies, J. T., Orrall, F. Q., Zirker, J. B., **The interpretation of total line intensities from optically thin gases. I. A general method,** *Solar Phys.* **22,** No. 2, 307-316 (Feb. 1972).

12925. Jefferies, J. T., Orrall, F. Q., Zirker, J. B., **The interpretation of total line intensities from optically thin gases. II. The coronal forbidden lines,** *Solar Phys.* **22,** No. 2, 317-326 (Feb. 1972).

12926. Jefferies, J. T., Orrall, F. Q., Zirker, J. B., **The interpretation of total line intensities from optically thin gases. III. Application to coronal forbidden line spectra,** *Solar Phys.* **22,** No. 2, 327-343 (Feb. 1972).

12927. Lightbody, J. W., Jr., **Electron scattering from one- and two-phonon vibrational states,** *Phys. Lett.* **38B,** No. 7, 475-479 (Apr. 3, 1972).

12928. Neuhausen, R., Lightbody, J. W., Jr., Fivozinsky, S. P., Penner, S., **Elastic electron scattering from Zn isotopes,** *Phys. Rev. C* **5,** No. 1, 124-128 (Jan. 1972).

12929. Bergstrom, J. C., Crannell, H., Kline, F. J., O'Brien, J. T., Lightbody, J. W., Jr., Fivozinsky, S. P., **Electroexcitation of the giant resonance of ^{13}C,** *Phys. Rev. C* **4,** No. 5, 1514-1532 (Nov. 1971).

12930. Pruitt, J. S., **Electron beam current monitoring system II,** *Nucl. Instrum. Methods* **100,** 433-443 (1972).

12931. Becker, D. A., **Trace analysis for platinum in glasses by neutron activation,** *Anal. Chim. Acta* **61,** 1-6 (1972).

12932. Pringle, W. C., Jr., **Microwave spectrum, vibration-rotation interaction, and ring puckering vibration in silacyclobutane and silacyclobutane-1,1-d_2,** *J. Chem. Phys.* **54,** No. 12, 4979-4988 (June 15, 1972).

12933. Peterson, R. L., **Longitudinal magnetoresistance in polar semiconductors: A displaced Maxwellian analysis,** *Phys. Rev. B* **5,** No. 10, 3994-4000 (May 15, 1972).

12934. Wiese, W. L., Kelleher, D. E., Paquette, D. R., **Detailed study of the Stark broadening of Balmer lines in a high-density plasma,** *Phys. Rev. A* **6,** No. 3, 1132-1153 (Sept. 1972).

12935. Moore, G. A., **Automatic methods for analysis of microstructures** (American Society for Testing and Materials Symp. on Stereology and Quantitative Metallography, Atlantic City, N.J., June 28, 1971), *Amer. Soc. Test. Mater. Spec. Tech. Publ.* **504,** pp. 59-80 (July 1972).

12936. Curl, R. F., Jr., Evenson, K. M., Wells, J. S., **Laser magnetic resonance spectrum of NO_2 at 337 μm and 311 μm,** *J. Chem. Phys.* **56,** No. 10, 5143-5151 (May 15, 1972).

12937. Billingsley, F. P., II, Krauss, M., **Coupled multiconfigurational self-consistent-field method for atomic dipole polarizabilities. I. Theory and application to carbon,** *Phys. Rev. A* **6,** No. 3, 855-865 (Sept. 1972).

12938. Jacox, M. E., Milligan, D. E., **Spectrum and structure of the O_3^- and O_4^- anions isolated in an argon matrix,** *Chem. Phys. Lett.* **14,** No. 4, 518-521 (June 15, 1972).

12939. Franklin, A. D., **Defect equilibria in alkaline earth fluorides containing rare earth ions,** *Tech. Rep.* **AFML-TR-72-31,** 48 pages (Air Force Materials Laboratory, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, Mar. 1972).

12940. Richmond, J. C., Kneissl, G. J., **Procedures for the precise determination of thermal radiation properties,** *Tech. Rep.* **AFML-TR-70-121,** 104 pages (Air Force Materials Laboratory, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio, Oct. 1970).

12941. Mountain, R. D., **Temperature dependence of depolarized scattered light near the critical point** (Proc. of the Colloquium on Light Scattering by Fluids, Paris, France, July 15-17, 1971), *J. Phys.* **33,** Supplement 2-3, C1-265–C1-268 (Feb.-Mar. 1972).

12942. McNesby, J. R., Braun, W., Ball, J., **Vacuum ultraviolet techniques in photochemistry,** Chapter 11 in *Creation and Detection of Excited State,* A. A. Lamola, ed., **1,** 503-586 (Marcel Dekker, Inc., New York, N.Y., 1971).

12943. Leasure, W. A., Jr., **Considerations in the development of measurement procedures for the regulation and monitoring of noise pollution** (Proc. 2d Urban Technology Conf. and Technical Display, San Francisco, Calif., July 24-26, 1972), *AIAA Paper No.* **72-622,** pp. 1-10 (American Institute of Aeronautics, New York, N.Y., 1972).

12944. Wood, L. A., Bullman, G. W., **Creep and other tensile properties of rubber crosslinked by dicumyl peroxide,** *J. Polym. Sci.:* Part A-2, **10,** 43-50 (Jan. 1972).

12945. Miller, A., McLaughlin, W. L., Lynggard, B., **A scanning spectrophotometer for reading thin-film dosimeters,** *Dan. At. Energy Comm. Risö Rep. No.* **M-1525,** pp. 1-21 (Aug. 1972).

12946. Beatty, R. W., **Efficiencies of microwave 2-ports from reflection coefficient measurements,** *IEEE Trans. Microwave Theory Tech.* **MTT-20,** No. 5, 343-344 (May 1972).

12947. Arthur, M. G., **A precision HF-noise power measurement system,** *ISA Trans.* **10,** No. 3, 264-268 (1971).

12948. Fong, J. T., Simmons, J. A., **A non-equilibrium thermodynamic theory of simple materials based on a single-integral entropic functional** (Proc. Int. Symp. on Foundations of Plasticity, Warsaw, Poland, Aug. 1972), *Arch. Mech. Stos.* **24,** No. 3, 363-372 (1972).

12949. Karp, S. S., Matteson, T. T., Maltese, M. D., **A search and rescue simulation model (SARSIM),** *Proc. Tokyo 1971 AICA Symp. on Simulation of Complex Systems, Tokyo, Japan, Sept. 3-7, 1971,* pp. I-4/1—I-4/8 (The Society of Analog Technique of Japan, Tokyo, Japan, Sept. 1971).

12950. Geltman, S., **Applications of pseudo-state expansions** (Proc. VII Int. Conf. on the Physics of Electronic and Atomic Collisions, Amsterdam, July 26, 1971), Chapter in *Physics of Electronic and Atomic Collisions,* T. R. Gover and F. J. de Heer, eds. pp. 216-231 (North-Holland Publ. Co., Amsterdam, The Netherlands, 1972).

12951. Coyne, J. J., Meshkov, S., **High-energy total cross sections and symmetry relations,** *Phys. Rev. D* **6,** No. 1, 399-402 (July 1, 1972).

12952. Hockey, B. J., **Observations by transmission electron microscopy on the subsurface damage produced in aluminum oxide by mechanical polishing and grinding** (Proc. British Ceramic Society Meeting, Textural Studies of Ceramics, London, England, Dec. 16-17, 1970), Paper No. 6 in *Textural Studies of Ceramics, Proceedings No.* **20,** pp. 95-115 (British Ceramic Society, Shelton, Stoke-on-Trent, Great Britain, June 1972).

**Publications with prices and SD Catalog numbers indicated may be purchased directly from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (foreign: one-fourth additional). NBS nonperiodical series are also available from the National Technical Information Service (NTIS), Springfield, Va., 22151. Reprints from outside journals and the NBS Journal of Research may often be obtained directly from the authors by title only.*

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Vol. 76B, Nos. 3 and 4, July–December 1972

INDEX TO VOLUME 76B

January to December 1972

A

Page

Abscissa 179
Abstracted groups as doubly transitive permutation groups 45
Area coordinates 145

B

Basic sets 153
Bernoulli 53
Bernoulli numbers 161
Binomial 53
Biorthogonal function 91
Boltzmann factors 193

C

Cannon sum of basic sets 153
Character 45
Checking parameters 61
Checks of the algebraic matrices for the configurations $(d+s)*p$ 61
Coin-tossing 53
Compound 205
Computer 53
Computer systems 109
Confidence limits 179
Confidence limits for the abscissa of intersection of two linear regressions 179
Cosets 45
Covariences of two sample rank sum statistics 51

D

$(d+s)*p$ 61
Difference equations 201
Distance coordinates 145
Distance coordinates with respect to a triangle of reference 145
Divisors 205

E

Efficiency 53
Efficient teachniques for unbiasing a Bernoulli generator 53
Electrostatic interaction 61
Euler-Maclaurin method 161
Euler transform 161
Euler's constant 161

F

Farag, M. A., Nassif, M., On regular sets of polynomials whose zeros lie in prescribed domains 153
Field 119
Filliben, J. J., McKinney, J. E., Confidence limits for the abscissa of intersection of two linear regressions 179
Functional equations 201

G

Page

Gamma-ray transport 91
Generator 53
Glass transition 179
Goldberg, K., Distance coordinates with respect to a triangle of reference 145
Goldman, A. J., Minimax selection, adjustment and disaggregation for incompletely specified univariate distributions 1
—, Product solutions and separation of variables 201
—, Scheduling a time-shared server to minimize aggregate delay 109

H

Helly's theorem 39
Horn, W. A., Three results for trees, using mathematical induction 39

I

Importance Sampling 193
Intercept 179
Intersection 179

L

Lechner, J. A., Efficient techniques for unbiasing a Bernoulli generator 53
Liang, J. J. Y., Tood, J., The Stieltjes constant 161
Linear 179
Linear programs 1
Lower semiblock matrices 153

M

Mann-Whitney-Wilcoxon statistics 51
Marcus, M., Underwood, E. E., A note on the multiplicative property of the Smith normal form 205
Mathematical induction 39
Mathematical models 1
Matrices 119
Matrix 205
Matrix representation 45
McCrackin, F. L., Weighting methods for Monte Carlo calculation of polymer configurations 193
McKinney, J. E., Filliben, J. J., Confidence limits for the abscissa of intersection of two linear regressions 179
Merris, R., Abstracted groups as doubly transitive permutation groups 45
Minimax-based selection, adjustment and disaggregation for incompletely specified univariate distributions 1
Minimax estimation 1
Minimum-length coverings 39
Moment methods 91
Monte Carlo 193
Multiple precision package 161

N

Page

Nassif, M., Farag, M. A., On regular sets of polynomials whose zeros lie in prescribed domains 153
neutron transport 91
Node-covering 39
Non-self-intersecting walks 193
Normal matrices with entries from an arbitrary field of characteristic ≠2 119
Normality 119
Normalizer 45
Note on the multiplicative property of the Smith normal form 205

O

On regular sets of polynomials whose zeros lie in prescribed domains 153
Operations Research 1
Optimization 109
Order of basic sets 153

P

Pearl, M. H., Penn, A. I., Normal matrices with entries from an arbitrary field of characteristic ≠ 2 119
Penn, A. I., Pearl, M. H., Normal matrices with entries from an arbitrary field of characteristic ≠ 2 119
Plane geometry 145
"Plural Series" approximations of functions 91
Polymer configurations 193
Polynomial representations 91
Probability distribution 1
Product solutions and separation of variables 201

R

Radiation penetration 91
Radical center 145
Random numbers 53
Random walks 193
Rank sum statistics 51
Regression 179
Relative phases 61
Rieman zeta function 161

Page

Roth, C., Checks of the algebraic matrices for the configurations $(d+s)^a p$ 61

S

Sample problem 51
Scheduling 109
Scheduling a time-shared server to minimize aggregate delay 109
Second order transition 179
Separation of variables 201
Solvable group 45
Spencer, L. V., "Plural Series" approximations of functions 91
Spin-orbit interaction 61
Statistics 51, 53, 179
Stieltjes constants 161

T

Three results for trees, using mathematical induction 39
Time-sharing 109
Todd, J., Liang, J. J. Y., The Stieltjes constants 161
Trees 39
Triangle of reference 145
Tryon, P. V., Covariances of two sample rank sum statistics 51

U

Underwood, E. E., Marcus, M., A note on the multiplicative property of the Smith normal form 205

V

Variance 179

W

Weighting methods for Monte Carlo calculation of polymer configurations 195

Z

Zeros of polynomials in regular sets 153

U. S. GOVERNMENT PRINTING OFFICE : 1973 O - 479-206

SJR-B DEC-74
UNIVERSITY MICROFILMS
JOY REED
300 N ZEEB RD
ANN ARBOR MI 48106

UNITED STA.
GOVERNMENT PRINTING
PUBLIC DOCUMENTS DEPARTME.
WASHINGTON, D.C. 20402

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

POSTAGE AND FEES PAID
U.S. GOVERNMENT PRINTING OFFICE

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences

76B, Nos. 3&4 July–December 1972

Contents

	Page
"Plural-series" approximations of functions. L. V. Spencer.	91
Scheduling a time-shared server to minimize aggregate delay. A. J. Goldman.	109
Normal matrices with entries from an arbitrary field of characteristic $\neq 2$. Martin H. Pearl and Alan I. Penn.	119
Distance coordinates with respect to a triangle of reference. Karl Goldberg.	145
On regular sets of polynomials whose zeros lie in prescribed domains. M. A. Farag and M. Nassif.	153
The Stieltjes constants. J. J. Y. Liang and John Todd.	161
Confidence limits for the abscissa of intersection of two linear regressions. James J. Filliben and John E. McKinney.	179
Weighting methods for Monte Carlo calculation of polymer configurations. Frank L. McCrackin.	193
Product solutions and separation of variables. A. J. Goldman.	201
A note on the multiplicative property of the Smith normal form. Marvin Marcus and Ernest E. Underwood.	205
Publications of the National Bureau of Standards.	207
Index to volume 76B, January to December 1972.	219

Library of Congress Catalog Card Number: 62–4414

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. (Order by SD Catalog No. C 13.22/sec.B:76B (3&4)). Price $2.25 (single copy). Subscription price: $9.00 a year; $2.25 additional for foreign mailing.

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1972